

University of Groningen

Inferring Diversification Rate Variation From Phylogenies With Fossils

Mitchell, Jonathan S; Etienne, Rampal S; Rabosky, Daniel L

Published in:
Systematic biology

DOI:
[10.1093/sysbio/syy035](https://doi.org/10.1093/sysbio/syy035)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Mitchell, J. S., Etienne, R. S., & Rabosky, D. L. (2019). Inferring Diversification Rate Variation From Phylogenies With Fossils. *Systematic biology*, 68(1), 1-18. <https://doi.org/10.1093/sysbio/syy035>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

For peer review only. Do not cite.

Inferring Diversification Rate Variation From Phylogenies With Fossils

Journal:	<i>Systematic Biology</i>
Manuscript ID	USYB-2017-048.R3
Manuscript Type:	Regular Manuscript
Date Submitted by the Author:	n/a
Complete List of Authors:	Mitchell, Jonathan Etienne, Rampal; University of Groningen, Community and Conservation Ecology Group Rabosky, Daniel; University of Michigan, Ecology and Evolutionary Biology
Keywords:	Macroevolution, Paleontology, Birth-death, Diversification, Extinction

SCHOLARONE™
Manuscripts

RH: ESTIMATING RATES USING FOSSILS AND PHYLOGENIES

Inferring Diversification Rate Variation From Phylogenies With Fossils

Jonathan S. Mitchell^{1, 2}, Rampal S. Etienne³, and Daniel L. Rabosky^{1, *}

¹ Museum of Zoology and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109

² Department of Biology, West Virginia University Institute of Technology, Beckley WV 25801

³ Groningen Institute for Evolutionary Life Sciences, University of Groningen, Box 11103, 9700 CC Groningen, The Netherlands

*Correspondence to be sent to: Museum of Zoology and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109

Email: drabosky@umich.edu

ABSTRACT

Time-calibrated phylogenies of living species have been widely used to study the tempo and mode of species diversification. However, it is increasingly clear that inferences about species diversification — extinction rates in particular — can be unreliable in the absence of paleontological data. We introduce a general framework based on the fossilized birth-death process for studying speciation-extinction dynamics on phylogenies of extant and extinct species. The model assumes that phylogenies can be modeled as a mixture of distinct evolutionary rate regimes and that a hierarchical Poisson process governs the number of such rate regimes across a tree. We implemented the model in BAMM, a computational framework that uses reversible jump Markov chain Monte Carlo to simulate a posterior distribution of macroevolutionary rate regimes conditional on the branching times and topology of a phylogeny. The implementation we describe can be applied to paleontological phylogenies, neontological phylogenies, and to phylogenies that include both extant and extinct taxa. We evaluate performance of the model on datasets simulated under a range of diversification scenarios. We find that speciation rates are reliably inferred in the absence of paleontological data. However, the inclusion of fossil observations substantially increases the accuracy of extinction rate estimates. We demonstrate that inferences are relatively robust to at least some violations of model assumptions, including heterogeneity in preservation rates and misspecification of the number of occurrences in paleontological datasets.

Keywords: macroevolution, fossils, speciation, extinction, comparative methods, BAMM

Variation in rates of speciation and extinction contribute to striking disparities in species richness among different clades of organisms. The study of macroevolutionary rates was pioneered in the fossil record, with the application of phenomenological birth-death models to patterns of diversity through time (Raup et al. 1973; Sepkoski 1978; Raup 1985). Phylogenies of extant species are widely used to estimate rates of both speciation and extinction (Nee et al. 1994; Nee 2006; Alfaro et al. 2009; Morlon 2014). However, the integration of paleontological and neontological data will likely provide more accurate insights into macroevolutionary dynamics than molecular phylogenies alone (Quental and Marshall 2010; Liow et al. 2010a; Stadler 2013; Rabosky 2016). Indeed, a recent simulation study has argued that macroevolutionary rates inferred using only a tree topology and fossil occurrence dates are more reliable than rates inferred with an extant-only tree and branch lengths (Didier et al. 2017).

Inferences on macroevolutionary patterns frequently differ between paleontological and molecular phylogenetic data (Hunt and Slater 2016). Fossil data routinely support high levels of extinction and large fluctuations in standing richness (e.g., Alroy 1999, 2008, 2010; Liow et al. 2010a; Quental and Marshall 2010; Sheets et al. 2016). However, extant-only analyses rarely find support for high extinction rates (Nee 2006; Purvis 2008, but see Etienne et al 2012), and numerous studies have addressed potential reasons for this discrepancy (Rabosky 2009, 2010, 2016; Morlon et al. 2011; Etienne & Rosindell 2012; Etienne et al. 2012; Rosenblum et al. 2012; Beaulieu and O'Meara 2015). Directly incorporating fossil data into phylogenetic rate estimation should facilitate more robust tests of the role of extinction in macroevolutionary dynamics (Sakamoto et al. 2016). An explicitly phylogenetic approach to macroevolutionary rates in the fossil record will provide insights into rate variation among lineages and will also facilitate the

study of diversification dynamics in clades that are too poorly preserved for subsampling-based approaches (e.g., Alroy 1999).

In this article, we describe a general framework for modeling complex speciation-extinction dynamics on phylogenetic trees that include both paleontological and neontological data. The model is based on the fossilized birth-death process (Stadler 2010; Didier et al. 2012, 2017; Heath et al. 2014) and includes parameters for speciation, extinction, and fossilization rates. The birth-death process for reconstructed phylogenetic trees (Nee et al. 1994) is a special case of the fossilized birth-death process where the preservation rate is equal to zero (Stadler 2010). As such, this model can be applied to paleontological phylogenies (extinct taxa only), neontological phylogenies of living species (extant taxa only), and phylogenies that include any combination of extant and extinct species.

The novel feature of our implementation, relative to previous work on diversification rates from phylogenies and fossils (Stadler 2010; Didier et al. 2012; Heath et al. 2014), is the ability to capture complex patterns of among-lineage variation in speciation and extinction rates. The model assumes that phylogenies are shaped by a countably finite set of distinct evolutionary rate regimes, and that the number of such regimes is governed by a hierarchical Poisson process. This model is an extension of the multiprocess diversification model currently implemented in the BAMM software package for macroevolutionary rate analysis (Rabosky 2014; Rabosky et al. 2017). We note that Sakamoto et al (2016) developed a flexible regression-based framework for estimating heterogeneity in diversification rates from fossil phylogenies, but their approach does not rely on a formal process-based model of diversification and preservation.

We first describe the analytical formulation of the model and state its assumptions. We then describe our implementation within the BAMM software framework and discuss its

relationship to previous BAMM versions (Rabosky et al. 2017). To evaluate the performance of the model, we simulated phylogenies under a range of diversification scenarios, and then subjected each simulated phylogeny to stochastic fossilization to generate a tree similar to what researchers would typically use (an “observed” phylogenetic tree with an incompletely fossilized history). We analyzed each phylogeny with BAMM and compared the BAMM-inferred rates and shift configurations to the known (simulated) values to assess the accuracy of parameter estimates. We find that the inclusion of even sparse paleontological data in phylogenies dominated by extant lineages improves the accuracy of both speciation and extinction rate estimates. This improvement is especially pronounced in estimation of extinction rates. Although our model makes several simplifying assumptions about the nature of fossilization, we demonstrate that rate inferences are reasonably robust to temporally-heterogeneity in preservation rates.

METHODS

Data

The data consist of a time-calibrated phylogenetic tree (which may include both extinct and extant lineages) and a count of the number of stratigraphically-distinct fossil occurrences associated with the tree. Each extinct lineage in the phylogeny should be represented by a branch that terminates at the time of the last occurrence for the taxon. Note that true extinction events are not observed directly, and the last occurrence time is expected to precede the true extinction time of a taxon. The number of fossil occurrences is best thought of as an estimate of the number of fossils that could be, in principle, assigned to lineages present in the tree. Fossil occurrences of indeterminate nature and that are potentially associated with lineages not present in the tree

(e.g, "*Theropoda, indeterminate*") are not accommodated in this modeling framework and are excluded from the count of occurrences; the likelihood calculations only apply to fossils that can be assigned explicitly to branches in the tree. Put another way, fossil occurrences should only be counted if the species-level taxon to which they are assigned is included in the tree. However, because the current implementation assumes homogeneity of preservation through time and among lineages, the calculations do not actually use stratigraphic and phylogenetic context for individual fossil observations. As explained below, this assumption of time-homogeneous preservation means that the likelihood of the phylogeny with fossil data is invariant with respect to the location of the fossils on the tree.

Likelihood of a Phylogeny with Fossils and Rate-Shifts

The fundamental operation in the analysis of diversification rate shifts on phylogenetic trees involves computing the likelihood of the phylogeny under a fixed configuration of lineage types. Each "type" of lineage is characterized by a distinct and potentially time-varying rate of speciation (λ) and extinction (μ). Two lineages i and j are said to be of the same type if $\lambda_i(t) = \lambda_j(t)$ and $\mu_i(t) = \mu_j(t)$ everywhere in time. Thus, for a given point in time t , lineages of the same type will have precisely the same distribution of progeny lineages at some time $t + \Delta t$ in the future. For the constant-rate birth-death process, all lineages are assumed to be of the same type. For the Binary State Speciation and Extinction (BiSSE) process (Maddison et al. 2007), phylogenies are assumed to comprise a mixture of two types of lineages, and the character states at the tips of the tree can be viewed as labels for lineage types. The transition points between types are unknown in the BiSSE process, and the likelihood of a given set of labeled types can be

computed by integrating over all possible transitions that could have given rise to the observed data.

In contrast to BiSSE, the calculations implemented in BAMM and other rate-shift models (Alfaro et al. 2009; Morlon et al. 2011; Etienne and Haegeman 2012) assume that we have mapped lineage types across all branches of the phylogeny and not merely to the tips. We define a "mapping" of rate parameters as an association between each point on a phylogeny (e.g., segment of a branch) and a particular lineage type. Additional steps may be used to optimize the number, location, and parameters of the mapped types, including maximum likelihood (Alfaro et al. 2009; Morlon et al. 2011; Etienne and Haegeman 2012).

Likelihood calculations for the fossilized birth-death process were described by Stadler (2010), and our general approach extends these calculations to a phylogeny with multiple types of lineages. The fossilized process differs from the reconstructed evolutionary process (Nee et al. 1994) because we must also account for the lineage fossilization rate ψ . The new calculations described below are implemented in BAMM v2.6.

The derivation of the likelihood for rate-shift models follows the logic described in Maddison et al's (2007) description of the BiSSE model. In a given infinitesimal interval of time Δt , the state-space for the process includes five events that can potentially change the probability of the data: a lineage may undergo a rate shift, become preserved as fossil, become extinct, speciate, or it may undergo none of these events. To compute the likelihood of the tree and a set of parameters (e.g., a mapping of rate shifts), we assume that no other rate shifts have occurred, thus dropping one of the possible events (the occurrence of an unmapped rate shift) from the state space for the process. As such, the likelihood is conditioned on the set of shifts that have been placed on the tree. This assumption of rate-shift models has been criticized in the recent

literature (Moore et al. 2016). However, it is unclear how we might accommodate unobserved
rate shifts, even in principle, as there is very little information in the data or elsewhere that can
yield information about the probability distribution of unobserved rate parameters. Most
importantly, we demonstrate that the simplifying assumptions that underlie the BAMM
calculations yield robust inference on evolutionary rates, even when simulated phylogenies used
for validation contain many such unobserved rate shifts.

We assume that preservation events are mutually exclusive with respect to speciation and
extinction events. Hence, a lineage cannot leave a recoverable fossil observation during the
precise moment of lineage splitting or lineage extinction. This interpretation of the preservation
rate ensures that the model we have implemented is mathematically identical to the model
described by Stadler (2010), for the special case where speciation and extinction rates do not
vary among lineages. This interpretation of preservation is also consistent with paleontological
practice: fossil observations are always assigned to single lineages. Even in the relatively high-
resolution marine microfossil record, paleontologists do not assume that the data capture the
precise instant of lineage splitting (Benton and Pearson 2001). In this article, we assume that
lineage preservation is a time-homogeneous process, but it is straightforward to relax this
assumption.

We measure time t backwards from some arbitrary starting time, t_0 . The start time, t_0 ,
need not correspond to the present. Let $D(t)$ be the probability density that a lineage that is extant
some time t before t_0 gives rise to the observed data at the present time, and let $E(t)$ be the
corresponding probability that a lineage that is extant at time t goes extinct, along with its
descendants, before time t_0 . Following Stadler (2010), the master equations governing the
likelihood of the data are

176

$$\frac{dD}{dt} = -(\lambda + \mu + \psi)D(t) + 2\lambda D(t)E(t) \tag{1}$$

178 and

$$\frac{dE}{dt} = \mu - (\lambda + \mu + \psi)E(t) + \lambda E(t)^2 \tag{2}$$

180 $E(t)$ describes the probability that an independent lineage originating at time t goes extinct, along
182 with all of its descendants, before some future time t_0 . One can also view this probability as the
probability of "non-observation" of a lineage and all of its descendants; a lineage and its
184 descendants might not be observed because they have gone extinct, and/or because we have
failed to sample them. To compute the likelihood of the phylogeny with mapped rate shifts, we
186 require solutions to these equations for arbitrary t and initial values t_0 , D_0 , and E_0 . For $E(t)$, the
solution is (Stadler 2010)

188

$$E(t) = \frac{\lambda + \mu + \psi + c_1 \frac{e^{-c_1 \Delta t} (1 - c_2) - (1 + c_2)}{e^{-c_1 \Delta t} (1 - c_2) + (1 + c_2)}}{2\lambda} \tag{3}$$

190 where

192

$$c_1 = \left| \sqrt{(\lambda - \mu - \psi)^2 + 4\lambda\psi} \right| \tag{4}$$

194 and

$$c_2 = -\frac{\lambda - \mu - \psi - 2\lambda(1 - E_0)}{c_1} \quad (5).$$

For $D(t, t_0)$, we have

$$D(t) = \frac{4D_0}{2(1 - c_2^2) + e^{-c_1\Delta t}(1 - c_2)^2 + e^{c_1\Delta t}(1 + c_2)^2} \quad (6)$$

which follows immediately from Stadler's (2010) Theorem 3.1, under the substitution $D_0 = \rho$. In the Appendix, we provide a derivation of $D(t)$ for arbitrary D_0 .

Calculations are performed by solving $D(t)$ and $E(t)$ recursively on the phylogeny from the tips to the root, combining $D(t)$ probabilities at internal nodes (discussed below) and continuing down the tree in a rootwards direction. For lineages that are extant, we have initial conditions at time $t_0 = 0$ of $D_0 = \rho$ and $E_0 = 1 - \rho$, where ρ is the probability that an extant lineage has been sampled (e.g., the sampling fraction; FitzJohn et al. 2009). In the absence of perfect preservation, the last fossil occurrence of a given taxon is not the true extinction time. For a lineage with a last occurrence at time t_i , we generally know that the lineage and all of its descendants went extinct sometime between t_i and $t_0 = 0$. Hence, for terminal lineages with non-extant tips, we initialize both E_0 and D_0 with $E(t_i | t_0 = 0)$, or the probability that a single lineage that is extant at time t_i will have zero observed descendants prior to and including the present day ($t = 0$). This initialization for D_0 accounts for the probability that a non-extant tip goes extinct, or is unsampled, on the time interval $(t_i, 0)$. At internal nodes, we combine the densities from the

right $D_R(t)$ and left $D_L(t)$ descendant lineages, such that $D(t)$ immediately rootwards of the node is computed as $D(t) = \lambda(t) D_R(t) D_L(t)$. The full likelihood must also account for all Z observed fossil occurrences, which is accomplished by multiplying the probability density of the tree by Ψ^Z .

There is one difference between the implementation of the model described here (implemented in BAMM v2.6) and the default calculation scheme implemented in the most recent previous version, BAMM v2.5. As discussed in Rabosky et al. (2017), there are several algorithms for computing the likelihood of a specified mapping of rate regimes on a phylogenetic tree. These algorithms, which we have termed "recompute" and "pass-up" (Rabosky et al 2017), differ in how the analytical $E(t)$ equation is initialized for calculations on individual branch segments. The precise sequence of calculations associated with these approaches is described in the supporting information (section S2.4) from Rabosky et al. (2017). Here, we expand BAMM to include the "recompute" algorithm for computing the likelihood, which was not implemented as the default calculation scheme in BAMM v2.5. The process under consideration in the present article can yield phylogenies that have gone extinct in their entirety (e.g., fossil taxa only), and we can thus avoid conditioning on survival of the process as well as unresolved theoretical problems associated with this conditioning. We thus perform all calculations using both the "recompute" and "pass-up" algorithms; we present "recompute" results in the main text, but the corresponding results using "pass-up" were virtually identical and are presented in the supplemental information.

As noted above, our use of these equations for modeling diversification dynamics implicitly conditions the likelihood on a finite and observed set of lineage types: the likelihood does not account for new types of lineages that may have evolved on lineages with no fossilized

or extant descendants (Moore et al 2016). This assumption applies to all methods that purport to compute the likelihood of a phylogeny under a specified mapping of rate shifts (Alfaro et al 2009, Morlon et al 2011, FitzJohn 2010, Etienne and Haegeman 2012). Indeed, this assumption applies indirectly to all diversification methods: if there are no (or very few) unobserved shifts, the likelihood computed by BAMM is approximately correct. If "reality" contains many such unobserved shifts, then the likelihood (and corresponding rate estimates) from BAMM will be biased, because rate shifts will essentially have served as an unaccommodated source of lineage extinction. But if unobserved rate shifts are sufficiently common (in real data) as to bias BAMM, then they are problematic for all other current methods and models as well. As pointed out by Rabosky et al (2017), the problem of unobserved rate shifts does not disappear simply because one assumes a theoretically complete model that ignores their existence or by sampling their parameters from a prior (assumed) distribution.

Implementation in BAMM

The model described above is implemented as an extension to the BAMM software (v2.6). BAMM v2.6 expands the basic diversification model implemented in earlier versions of BAMM by incorporating parameters that specify the total observation time of the tree ("observationTime", T_{obs} : time from root when the tree is observed; this parameter is equal to the root age for trees with extant lineages), the total number of observed fossil occurrences within the clade analyzed ("numberOccurrences"), and parameters that control the prior distribution of, and update frequency for, the preservation rate Ψ ("updatePreservationRate", "preservationRatePrior", and "preservationRateInit").

T_{obs} can have a large impact on the overall model fit, and warrants additional explanation.

264 The model assumes that non-extant tip lineages can go extinct anywhere on the time interval that
extends from their last occurrence to T_{obs} . Consider the scenario where we are modeling
266 diversification dynamics in an extinct clade of dinosaurs that originated 200 million years before
the present (Ma) and where the last occurrence of that clade is 90 Ma. If we set T_{obs} to equal the
268 present day (0 Ma), the model will allow the possibility that the last dinosaur from the clade went
extinct sometime between 90 and 0 Ma. In practice, we suggest choosing the observationTime as
270 the earliest possible time that all last-occurrence taxa are believed to have gone extinct. In the
dinosaur example, a reasonable setting for T_{obs} would be the KPg boundary at 66 Ma, by which
272 time all non-avian dinosaur lineages had become extinct.

The number of occurrences (parameter “numberOccurrences” in BAMM) represents the
274 total number of fossil occurrences that can be assigned to observed lineages in the clade under
consideration and is not simply the number of extinct tips in the phylogeny. It is necessarily the
276 case that a phylogeny with k extinct tips has at least k occurrences: one for the last occurrence of
each extinct lineage; however, there can be multiple occurrences for each extinct and extant
278 lineage. There are several possible interpretations of a fossil occurrence in the BAMM
framework; we address this issue at length in the Discussion. Occurrences associated with fossil
280 taxa that are not present in the tree should not be used, nor should taxonomically indeterminate
material (e.g., fragmentary fossil material that cannot be taxonomically identified below the
282 genus level). In general, we consider an “occurrence” to represent stratigraphically-unique
species-level observations. Hence, a locality with 1,000 individuals of a single taxon
284 representing a single depositional event (e.g., a mass burial event) would constitute a single

occurrence. In practice, we suggest analyzing a dataset with several different values for the
occurrence count to ensure that results are robust (see Discussion).

Description of Tree Simulation

We developed software to simulate phylogenies under a Poisson process that includes shifts to new macroevolutionary rate regimes (<https://github.com/macroevolution/simtree>). Each simulation was initialized with two lineages at time $t = 0$ with λ and μ drawn from exponential distributions with means of 0.2 and 0.18, respectively. We assumed that shifts to new rate regimes occurred with rate $\lambda = 0.02$. Waiting times between events (speciation, extinction, rate shift) were drawn from an exponential distribution with a rate equal to the sum of the rate parameters ($\lambda + \mu + \lambda$). For rate shift events, new values of λ and μ were drawn from the same exponential distributions as the root regime. This procedure simulates a phylogeny with stochastic rate shifts; within a given rate regime, λ and μ were constrained to be constant in time. Simulations were run for 100 time units, and trees with more than 5000 or fewer than 50 lineages at $t = 100$ were discarded, resulting in a total of 200 trees for analysis. In addition, each tree was constrained to have at least one rate shift in the true, generating history. This simulation procedure can generate phylogenies with unobserved rate shifts when shifts occur in unsampled extinct clades (see below). Conversely, the true root regime may be unobserved and all observed tips may fall within one of the derived regimes. As discussed in the section below on fossilization, any rate shift that leads to an extinct clade with no fossil observations will be unobserved, thus enabling us to assess whether Moore et al's (2016) concerns about unobserved rate shifts have consequences for inference in practice.

308 *Description of Fossilization Procedure*

We created pseudo-fossilized trees by stochastically placing fossils along the branches of
310 each simulated tree (Fig. 1) using per-lineage preservation rates (ψ) of 0, 0.01, 0.1 or 1.0.
Speciation rates in our simulations ranged from 0.003 to 1.7 lineages per unit time, with 99% of
312 the regimes having a speciation rate above 0.015; 80% exceeded 0.1. As such, speciation events
were generally more frequent than fossilization events, except under the high preservation ($\psi =$
314 1.0) simulation scenario. All extant tips, and tips with at least one fossil occurrence, were
retained while tips (and clades) lacking fossil occurrences were dropped. In empirical datasets,
316 the true extinction time of an extinct lineage is unknown, and the time of the last (most recent)
fossil occurrence is used. For comparability with empirical data from the fossil record, extinct
318 tips with fossils were truncated to the time of the most recent fossil occurrence (e.g., the branch
ends at the last occurrence, and any subsequent history for the lineage is necessarily unobserved).
320 We refer to these pseudo-fossilized trees as 'degraded', because the stochastic fossilization
history necessarily discards information about the true lineage history.

322 As pointed out by Moore et al (2016), the model implemented in BAMM assumes that
rate shifts do not occur on unobserved branches. However, our simulation and fossilization
324 procedure allows rate shifts to occur on extinct clades with no fossil observations. Hence, our
fossilization procedure frequently leads to unobserved rate regimes and enables us to test the
326 practical significance of this core BAMM assumption.

328 *BAMM Analyses*

We approximated the posterior distribution of rate shift configurations for each degraded
330 tree by running a Markov chain Monte Carlo (MCMC) simulation in BAMM v2.6 for 50 million

generations, with the first 5 million discarded as burnin. Each tree was analyzed assuming a single, time-constant rate of preservation. The model prior (mean of the geometric prior on the number of shifts) was set to 200 for all runs (after Mitchell and Rabosky 2016), while the rates of the exponential priors on λ and μ were scaled using the `setBAMMpriors()` function in BAMMtools (see Supplementary Material for details). Mitchell and Rabosky (2016) found that inferences on the number of shifts were generally robust to the choice of model prior, but that use of a liberal model prior facilitated convergence of the MCMC simulation. We performed each analysis using both "recompute" and "pass-up" algorithms for computing the likelihood, as described by Rabosky et al (2017). All computer code, data files, and result files are available through the Dryad digital data repository (doi:10.5061/dryad.p009h).

Performance Assessment

To assess the performance of our model and BAMM implementation, we quantified four fundamental attributes of interest: 1) the estimated preservation rate, 2) the speciation and extinction estimates inferred for each (true) rate regime, 3) the inferred number of macroevolutionary regimes on the tree, and 4) the location of inferred rate shifts. We compared the posterior distribution of the estimated preservation rate ψ for each degraded tree to the generating value. For diversification rates we found the mean inferred speciation and extinction rates for every true regime with at least two tips (both extinct and extant), and compared these mean values to the true values of speciation (λ), extinction (μ) and relative extinction (μ / λ). We predicted that BAMM would estimate rates more accurately for regimes with larger numbers of taxa; we thus assessed BAMM's accuracy with explicit reference to the number of tips in each regime. This allowed us to assess how well BAMM estimates rates for regimes that differ in size.

We determined the best-supported number of shifts for each tree by using stepwise model selection with Bayes factors (Mitchell and Rabosky 2016). Using the posterior distribution of shifts for each tree, we compared the lowest complexity model (e.g., smallest number of shifts) to the next-most complex model. If the more complex model was supported by Bayes factor evidence of at least 20 relative to the simpler one, it was provisionally accepted and the process repeated with models of increasing complexity until the Bayes factor threshold of 20 was no longer exceeded. We then compared the best-supported number of shifts on the tree to the true number of preserved shifts (see Fig. 1).

Finally, we looked at how accurately BAMM inferred the locations of shifts along the tree. Location accuracy is difficult to quantify, as many simple metrics have undesirable properties. For example, the distance along the branches between a true and inferred shift is difficult to interpret unless the number of inferred shifts is exactly equal to the number of true shifts. Branch lengths further confound location accuracy as shifts along shorter branches are expected to be more “smeared” across adjacent branches relative to shifts on longer branches (see extensive discussion of this topic in the supplement to Mitchell and Rabosky, 2016). We thus assessed location accuracy using several complementary metrics.

The first metric we used for location accuracy was the marginal posterior probability of at least one rate shift on all branches that included a (true) rate shift. We assessed these marginal probabilities with respect to both the magnitude of the shift (e.g., size of the rate increase) as well as the number of tips descending from that shift, with the expectation that shifts with many descendent tips and/or with a large change in rates would have higher marginal posterior probabilities. As pointed out by Rabosky et al (2014), focusing on these marginal probabilities can be misleading if the evidence for a rate shift is “smeared” across a set of adjacent branches.

As an alternative metric of location accuracy, we used the method described by Mitchell and Rabosky (2016) to assess how well BAMM partitions tips into their correct rate classes. This approach can be motivated by noting that any placement of k rate shifts on a phylogeny partitions a tree into at most $k + 1$ subsets of tips with distinct rate regimes ($k + 1$, not simply k , due to the presence of a root regime that is distinct from the rate shifts). For a phylogeny with N taxa and a known (true) mapping of diversification regimes, there is an $N \times N$ matrix where each element $C_{i,k}$ describes whether the i 'th and k 'th taxa are assigned to the same or different rate regimes (the set of tips partitioned into a particular rate regime was termed a “macroevolutionary cohort” in Rabosky et al. 2014). Elements of the matrix $C_{i,k}$ take a value of 1 (if taxa i and k are in the same rate regime), or 0 (if i and k are in different rate regimes). Following BAMM analysis, we obtain a matrix D where $D_{i,k}$ represents the posterior probability that taxa i and k are assigned to the same rate regime. To assess shift location accuracy, we computed one minus the average of the absolute value of the difference between the true pairwise partitioning matrix C and the BAMM-estimated matrix D , or

$$1 - \frac{2}{N(N-1)} \sum_{k=2}^N \sum_{i=1}^{k-1} |C_{i,k} - D_{i,k}| \quad (7)$$

An overall value of 1.0 can only be obtained if all pairs of taxa are correctly partitioned in the correct configuration of rate regimes in 100% of samples from the posterior. This metric of location accuracy will penalize shift configurations that include too many shifts (by separating taxa that should be in the same regime) and that include too few shifts (by uniting taxa that should be in separate regimes). However, the topology of the tree may make certain partitions more likely by chance alone. To quantify this baseline accuracy, we computed the partitioning

accuracy that would be expected by chance if shifts were distributed randomly throughout the tree under a uniform prior with respect to the total tree length. Under a uniform prior, the probability of a shift on a given branch b_i is simply b_i / T , where b_i is the length of the i 'th branch and T is the tree length. We conditioned the randomization on the simulated shift count, such that each randomized sample contained exactly the same number of shifts as the corresponding sample from the posterior. After randomizing the placements of shifts from all samples in the posterior, we computed the accuracy of random partitions for each sampled generation using the same pairwise distance matrix approach described above.

Sensitivity of the Method to Model Violations

We tested sensitivity of the framework described here to four qualitatively distinct violations of the assumptions of the underlying inference model. These violations included: (i) violation of time-homogeneous preservation potential; (ii) misspecification of the true number of fossil occurrences in the data; (iii) presence of a mass extinction in the data; and (iv) diversity-dependent diversification. In real datasets, preservation rates can vary substantially even over relatively short timescales (Foote and Raup 1996; Holland 2016). We violated the assumption of time-constant preservation by applying a “stage-specific” preservation pattern to the simulated phylogenies. We use "stage" in this context solely to denote an arbitrary temporal span with a potentially distinct preservation rate. We divided each simulated tree into ten equally-sized temporal bins (each ten time units long, mirroring the 10 myr bins used in some large-scale paleodiversity studies), where each bin was assigned an independent preservation rate drawn from a uniform (0.01, 0.9) distribution. We then simulated fossilization histories using these preservation rates and pruned unobserved lineage histories as described above to generate a set

of degraded trees. We analyzed these trees using BAMM v2.6, which assumes time-
homogeneous preservation, and we tested the accuracy of inference using the same metrics
described above. Given that the simulated trees involve rate shifts, this simulation approach
allows clades to radiate in periods of higher or lower preservation, and to persist into subsequent
time bins with distinct and decoupled preservation parameters.

Our implementation assumes that users can specify a meaningful number of fossil
occurrences for the clade. Fossil occurrences are not placed at specific points on the tree, but
rather inform the model about the total number of fossil occurrences associated with the focal
clade. The current implementation thus assumes fossil occurrences (with the exception of last
occurrences) are distributed across the tree under a homogeneous Poisson process, and estimates
a tree-wide rate accordingly. However, for real datasets, the “true” number of occurrences is
likely impossible to know with precision (even in principle; see Discussion). To assess how
misspecification of the number of fossil occurrences may bias inference, we simulated 100
constant rate phylogenies, with higher rates than the variable-rate tree simulations (mean $\lambda = 0.9$,
mean $\mu = 0.45$) and imposed a fossilization history with ψ of 0.1. Given the stochastic nature of
the fossilization simulation, the fossilized trees varied in the number of extinct lineages observed.
Note that because an extinct tip is only observed if at least one fossil occurs on that branch, each
fossilized tree with (observed) extinct lineages has a minimum number of fossil occurrences
equal to the number of (observed) extinct tips (F_{MIN}). We then calculated the difference between
this per-tree minimum and the true number ($F_{\text{TRUE}} - F_{\text{MIN}} = F_{\text{GAP}}$), and analyzed the tree using
BAMM v2.6 with the observed number of occurrences set to: (i) F_{MIN} , (ii) $F_{\text{MIN}} + 0.25 * F_{\text{GAP}}$,
(iii) $F_{\text{MIN}} + 0.5 * F_{\text{GAP}}$, (iv) F_{TRUE} , (v) $F_{\text{TRUE}} + F_{\text{GAP}}$, and (vi) $10 * F_{\text{TRUE}}$. Overestimation of the
true number of fossils may seem unlikely, but this could arise in practice due to taxonomic error

(e.g., incorrect assignment of fossil observations to a specific taxon) or to the inclusion of indeterminate fossil material in the analysis. Scaling the degree of misspecification to the true and minimum number of fossils created roughly comparable degrees of misspecification across trees with large differences in the number of observed extinct lineages.

We estimated the bias for each of these misspecified fossil occurrence values by dividing the inferred speciation and extinction rates across the entire tree (as inferred using the `getCladeRates` function from BAMMtools) by the true values used to simulate the tree. Our discussion below pertains to bias in rates resulting from misspecification of the number of occurrences, not bias due a particular magnitude of preservation (e.g., "low" or "high" preservation). We predicted that overestimates of fossil occurrences should lead to underestimates of evolutionary rates. Essentially, as the number of fossil occurrences is increased, the inferred preservation rate will be higher; this increased preservation rate, in turn, will decrease the amount of unobserved evolutionary history that is assumed to have occurred. By reducing the inferred unobserved history through artificial inflation of the preservation rate, we underestimate the number of events that have occurred (speciation or extinction), and the corresponding rate estimates are biased downwards. When the observed number of fossil occurrences is lower than the true number, however, we will tend to overestimate both the unobserved evolutionary history and the corresponding diversification rates.

We also tested the impact of diversification histories that violate the assumptions of the BAMM framework using two scenarios. First, we asked how the recent extinction of a large clade may bias the estimates from BAMM. To simulate this scenario, we chose a single, large, extant clade on each simulated phylogeny and pruned each extant branch back to the time of the most recent fossil occurrence (or dropped it entirely, if it was never observed in the fossil record).

For the second scenario, we simulated trees under a diversity-dependent process, and inferred rates from that tree both with and without fossil data. Further details on these scenarios are presented in the Supplementary Materials (Section S8).

Empirical Analysis

For comparison purposes, we include an empirical analysis using BAMM of a recently published 138-tip phylogeny of horses (Cantalapiedra et al. 2017). We time-scaled the horse tree using the same approach as Cantalapiedra et al. (2017), which tends to homogenize branches by assuming all branch lengths are pulled from a single distribution defined by a speciation, extinction and preservation rate (see Bapst, 2013; and the Supplementary Materials). We compared the BAMM results to those obtained using PyRate (Silvestro et al. 2014), a Bayesian inference framework that allows estimation of diversification parameters and rate shifts from the occurrence data alone. We downloaded 1495 fossil occurrences (see Discussion) for the 138 species in the phylogeny from the Paleobiology Database (Alroy et al. 2001) using the R interface ‘paleobioDB’ (Varela et al. 2015). We simulated the posterior distribution of rate shift configurations using BAMM with 20,000,000 generations of MCMC sampling. The PyRate analysis used 1,000,000 generations of MCMC sampling.

RESULTS

Simulated trees varied from 50 to 4578 tips (mean = 770), and degrading the trees reduced the observed number of lineages to an average size of 589, 616 and 698 tips for the lowest to highest preservation rates. Because multiple fossilization histories were imposed on each fixed evolutionary history, the number of fossil occurrences scales directly with the

492 preservation rate (mean numbers of fossils of 25.4, 255.6 and 2549.9 for the lowest to highest
preservation rates). After 50 million generations, the MCMC chains for at least 95% of datasets
494 in each fossilization scheme had reached our target convergence criterion, with effective sample
sizes for the number of shifts at over 200. MCMC simulations that failed to meet this
496 convergence criterion are nevertheless included in all following analyses to avoid biasing the
subsample by including only trees that showed good MCMC performance. Results based on all
498 trees, even those that failed to achieve the desired effective sample sizes, are nearly identical to
the results presented below and are shown in the supplementary material.

500 For each rate regime in each simulated tree, we compared the estimated rates with the
true rates in the generating process. Estimated speciation rates were reasonably well correlated
502 with true rates across all preservation levels. Across all rate regimes and preservation scenarios
the correlation between true and estimated speciation rates was 0.49, although 70.9% of these
504 regimes contained fewer than five tips. It is unlikely that BAMM would have had sufficient
power to infer such small regimes (Rabosky et al. 2017), and estimated rates for most will tend
506 to equal the rate of the parent process. For regimes with five or more tips, the correlation
increases to 0.77 and 0.86 for the extant-only and high-preservation ($\psi = 1.0$) scenarios, with all
508 other preservation scenarios falling between these values (Fig. 2, left column). For regimes with
20 or more tips, these correlations rise to 0.90 and 0.97, respectively (Fig. 3, left column). The
510 regression slopes are close to the expected value of one, but do not show a consistent increase
with preservation (extant-only: $\text{slope}_{5+} = 0.91 \pm 0.01$ s.e. and $\text{slope}_{20+} = 1.01 \pm 0.01$ s.e. for regimes
512 with 5+ tips and regimes with 20+ tips, respectively; $\psi = 0.01$: $\text{slope}_{5+} = 0.90 \pm 0.01$ s.e. and
 $\text{slope}_{20+} = 0.99 \pm 0.01$ s.e.; $\psi = 0.1$: $\text{slope}_{5+} = 0.86 \pm 0.01$ s.e. and $\text{slope}_{20+} = 0.94 \pm 0.01$ s.e.; $\psi = 1$:
514 $\text{slope}_{5+} = 0.88 \pm 0.01$ s.e. and $\text{slope}_{20+} = 0.96 \pm 0.01$ s.e.). Speciation rate correlations improve as

regimes with an increasing minimum number of tips are considered, regardless of preservation
 516 rate (Fig. 4).

Estimates of the extinction rate (μ) are substantially influenced by the preservation rate.

518 Across all regimes and preservation rates, the mean correlation between true and estimated
 regime-specific rates was 0.14. As for speciation, however, these correlations improve
 520 substantially when we exclude the 70.9% of rate regimes that included fewer than five tips. For
 regimes with five or more tips, correlations ranged from 0.30 for extant-only trees to 0.72 for the
 522 $\psi = 1.0$ scenario (Fig. 2). With 20 or more tips, these correlations increase to 0.49 and 0.91,
 respectively (Fig. 3). As with speciation, the slopes do not vary consistently with preservation,
 524 although the error decreases with increasing preservation (extant-only: $\text{slope}_{\text{all}} = 0.54 \pm 0.02$ and
 $\text{slope}_{20+} = 1.20 \pm 0.05$ for all and regimes with 20+ tips; $\psi = 0.01$: $\text{slope}_{\text{all}} = 0.47 \pm 0.02$ and
 526 $\text{slope}_{20+} = 0.95 \pm 0.05$; $\psi = 0.1$: $\text{slope}_{\text{all}} = 0.38 \pm 0.01$ and $\text{slope}_{20+} = 0.78 \pm 0.02$; $\psi = 1$: $\text{slope}_{\text{all}} =$
 0.50 ± 0.01 and $\text{slope}_{20+} = 0.95 \pm 0.01$).

528 The relative extinction rates (μ / λ) also improved with increasing preservation, with the
 estimates from extant-only simulations only weakly correlated with the true relative extinction
 530 level ($r = 0.41$ for 5+ tips, 0.53 for 20+ tips) and again increasing with increased preservation (r
 $= 0.67$ for regimes with 5+ tips; 0.87 for regimes with >20 tips; Fig. 2 and 3; right column). As
 532 for speciation, extinction rate correlations improve as we consider subsets of regimes with
 increasing numbers of tips, but accuracy of rate estimates is heavily influenced by the
 534 preservation rate (Fig. 4b). Preservation rates were estimated with high accuracy, with tight
 distributions around the generating values at $\psi = 0.01$, 0.1, and 1.0 (the 5 and 95% quantiles of
 536 the posterior estimate for $\psi = 0.01$ are $\psi_5 = 0.007$ and $\psi_{95} = 0.015$; $\psi_5 = 0.09$ and $\psi_{95} = 0.12$ for
 $\psi = 0.1$; $\psi_5 = 0.96$ and $\psi_{95} = 1.04$ for $\psi = 1$; Fig 5).

As has been shown previously (Rabosky 2014), BAMM tends to be relatively conservative in the number of regimes on the tree even with the high prior on the number of shifts used in our analyses (Fig. 6). Using stepwise model selection with a Bayes factor threshold of 20, we identified spurious shifts in fewer than 0.5% of trees (Fig. 6). We tabulated the marginal posterior probability of a rate shift for each branch on which a (true) rate shift occurred. These marginal probabilities varied substantially but generally increased with the number of descendent tips across all preservation regimes (Fig. 7a). The marginal probability of a shift along a particular branch increases with both the magnitude of the rate increase and with the number of descendant lineages (Fig. 7b). Thus, BAMM substantially underestimates the number of rate shifts (Fig. 6), but shifts that are not detected tend to be those leading to small clades and/or those that represent small changes in evolutionary rates (Fig. 7). Tip partitioning accuracy in BAMM v2.6 was high overall, and did not vary systematically with preservation rate (Fig. 7c).

The apparent conservatism of BAMM with respect to the true number of shifts need not be a problem: we have previously shown that, under the Poisson process, the vast majority of rate shifts lead to small rate regimes (< 5 tips), and that shift regimes of this size are unlikely to be inferred by any method (Rabosky et al 2017). Because the regimes are small (e.g., contain minimal data), their likelihood is similar across a broad range of parameter values, and there is insufficient information in the data to infer their existence. As pointed out by one reviewer, our results also raise a paradox: how can the BAMM-estimated rates perform well, despite the massive underestimate of the number of shifts? The solution to this paradox lies in the fact that small rate shifts are missed by BAMM, but dominant rate classes across the tree are captured by BAMM with reasonable accuracy. Hence, BAMM might correctly infer only 3 of 15 rate regimes in a given phylogeny, but those three inferred regimes can govern the majority of

branches in the tree (e.g., 12 shifts leading to one or two taxa, but the other three shifts all
leading to 50 or more taxa).

Sensitivity of the Method to Model Violations

We found that the implementation performed well even when assumptions of
homogeneous fossilization rates through time were violated. After imposing temporally-varying
preservation rates on the data, we found that diversification rates were still estimated with high
accuracy (Fig. 8). The estimates of both speciation and relative extinction (Fig. 8a,b) were
reasonably correlated with the generating values. We also found very few instances where the
number of inferred shifts was higher than the generating model (Fig. 8c). Consistent with the
time-homogenous preservation simulations, 78% of the tips were accurately partitioned among
the rate regimes (Fig. 8d).

Underestimating the number of fossil occurrences inflates the speciation and extinction
rates, while overestimating the number of occurrences deflates them (Fig. 9). Individual
estimates of λ may be off by as much as 50% when misspecification is exceptionally high (Fig.
9a), while μ may vary by substantially more (Fig. 9b). However, in aggregate across all 100 trees,
the median error at each level of occurrence misspecification is fairly low. Encouragingly, the
correlations of tree-wide rate estimates for these constant rate trees are high even when the fossil
occurrence numbers are off by a factor of ten (correlation for speciation: 0.95 to 0.89; correlation
for extinction: 0.93 to 0.89). This suggests that, if the number of fossil occurrences is known
only with large uncertainty, the relative rates across the phylogeny are likely to be more accurate
than the absolute values of the rates (e.g., shifts and temporal heterogeneity may still be detected
reliably, even if the rate magnitude is biased).

We explored BAMM's performance under two scenarios where the diversification process in the simulation model violated the assumptions of the inference model. For the diversity-dependent simulations, BAMM performed much better with the inclusion of fossil information (Fig. S11): both speciation and extinction rates were estimated with much greater accuracy when fossils were included, relative to the extant-only analysis. For the mass extinction scenario, we observed consistent biases in rate estimates when fossil and extant data were combined, and relative extinction rates were consistently biased upwards (Fig. S10).

Unobserved shifts seem to have a negligible impact on rate estimation. Neither the slope nor correlation coefficient of the extinction and speciation rates for branches within trees varies with the proportion of unobserved rate shifts (Fig. 10), nor with the absolute number of unobserved shifts (see Supplementary Material). In contrast to the regime-specific assessments presented above, within-tree measures are difficult to interpret under rate-shift models like BAMM. If a method does not infer rate heterogeneity within the tree, the metrics (especially the slope) are expected to perform poorly. Rabosky et al. (2017) discusses this issue in detail, documenting how the metrics may be misleading even when rates on almost every branch are estimated with high accuracy. Regardless, there is no obvious trend between the proportion of unobserved shifts within a tree and the overall metrics of performance (Fig. 10).

Empirical Analysis

Our empirical analysis of the horse phylogeny yielded evidence for a major shift in rates leading up to the extant genus *Equus*, and also that speciation and extinction rates are approximately equal across the tree and through time (Fig. 11). Net diversification was found to be near-zero across the horse phylogeny (mean net div. = -0.002; see Fig. 11), but lineage

turnover rates were substantial (mean turnover = 0.7). This result is concordant with long-standing hypotheses about the history of life, where speciation is only slightly elevated relative to extinction, and only in some clades (e.g., Raup 1994). However, it is worth noting that the BAMM estimates assume that rates have been constant through time within rate regimes, and it is unlikely that this assumption is strictly true in practice. It is possible, for example, that the near-zero net diversification rates estimated for equids represent the long-term average of a clade that experienced a pronounced phase of positive net diversification followed by negative net diversification rates that led to extinction of many lineages (Quental and Marshall 2013; Cantalapiedra et al. 2017). BAMM's assumption of time-homogeneous diversification means, in practice, that temporal rate variation will be smeared across the duration of a particular rate regime. We caution users of BAMM from overinterpreting rates at specific points in time if time-constancy of rates is assumed within rate regimes.

PyRate likewise infers an increase in extinction through time, but differs from the BAMM results in several ways. Extinction as estimated with PyRate was substantially lower than speciation during the earliest portion of horse evolution, but speciation dropped substantially at roughly 15Ma and rebounded slightly at 6Ma. The estimates of speciation and extinction from PyRate differ substantially from those estimated by BAMM, with greater volatility in the PyRate estimates of net diversification; BAMM inferred a more consistent near-zero level of net diversification. Notably, the turnover rates estimated by BAMM are much higher than those from PyRate. Both methods estimate high preservation rates, although as speciation, extinction, and preservation are estimated jointly, the two estimates are different, with BAMM estimating a higher rate ($\psi = 3.69 \pm 0.10$) than PyRate ($\psi = 2.27 \pm 0.07$).

630 DISCUSSION

We have described an extension of the BAMM framework for modeling
632 macroevolutionary dynamics that leverages recent advances in modeling the fossilized birth-
death process (Stadler 2010; Ronquist et al. 2012; Heath et al. 2014; Didier et al 2012, 2017) to
634 estimate rates of speciation, extinction, and preservation from phylogenies that include
observations of fossil (non-extant) lineages. We found that BAMM was able to infer
636 diversification rates, preservation rates, and the number and location of rate shifts with
reasonable accuracy across a range of fossilization scenarios. The inference model implemented
638 in BAMM is robust to at least some violations of its assumptions, although we necessarily
explored a limited subset of many possible ways that model assumptions can be violated. Given
640 debates over the inference of extinction from molecular phylogenies (Rabosky 2010, 2016;
Beaulieu and O'Meara 2015), it is noteworthy that inclusion of even a small number of fossils
642 (<10% of the total tips) substantially improves the estimation of both the absolute and relative
extinction rates in comparison to analyses of the same data restricted to extant tips only.
644 BAMM's likelihood is conditioned on the absence of unobserved rate shifts, an assumption that
has been criticized in the recent literature (Moore et al 2016). However, we find no evidence that
646 these unknown quantities affect inference (Fig. 10). Our results are thus consistent with Rabosky
et al's (2017) demonstration that unobserved rate shifts are unimportant for empirically relevant
648 parameterizations of the diversification process.

Given the preservational biases inherent to the fossil record, our results showing that
650 BAMM is at least somewhat robust to temporal heterogeneity in preservation rates and
uncertainty in the number of fossil occurrences are encouraging (Fig. 8, 9). Changes in the
652 preservation rate through time are likely to be more extreme in real datasets than considered by

our study. However, as long as the average preservation through time is comparable to values
used in our simulations, we expect that the inclusion of fossil data will improve estimates of
diversification parameters. Incorrect specification of the number of fossil occurrences biases the
magnitude of rate estimates, but the true and estimated rates remain highly correlated even with
substantial misspecification of the true number of fossils. This suggests that relative rates within
a phylogeny may largely be robust to misspecification of occurrence counts, even if the absolute
values of those rates are biased. Nonetheless, assumptions about the nature of fossil preservation
and lineage comparability between neontological and paleontological datatypes should be treated
with caution in future empirical work.

In the case of clade-wide mass extinction (Fig. S10), the bias in BAMM's extinction
estimates is easily understood. BAMM has no ability to accommodate mass extinctions directly,
but it is nonetheless forced to model a situation where a large number of taxa became extinct at a
particular point in time. Hence, the background extinction rate for the focal clade must increase
to accommodate the extinction of many taxa, even though the clade may have had low extinction
throughout much of its history. In effect, BAMM smears the effects of the mass extinction across
earlier periods of time when background extinction was lower.

Because the current implementation of the model does not accommodate time-varying
diversification rates, it is possible that resulting time-smeared rate estimates will be an
inadequate description of the true diversification process. Similarly, for any clades that have
gone extinct in their entirety, net diversification rates will generally be estimated as near zero
throughout the history of the clade. This potential for time-averaging of rates is a substantial
concern for interpretation that must be considered carefully by users of the method, at least until

extensions have been developed that can better accommodate time-varying diversification
processes.

Sampling Biases

Paleontologists have long recognized that apparent long-term trends in
macroevolutionary dynamics can result from biases in the rock record, and in particular the
temporal variation in the total amount of sedimentary rock (Raup 1972, 1976; Sepkoski et al.
1981; Foote and Raup 1996; Peters and Foote 2001; Peters 2006; Holland 2016). Even if total
species richness is constant through time, intervals with greater amounts of exposed sedimentary
rock volume, or with the same amount of rock distributed across a greater geographic extent, will
generally appear to be more diverse (Raup 1972). A general trend of increasing preservation
potential through time has long been documented (e.g., Raup 1976), and if this trend were
universally monotonic it would be relatively easy to model. However, preservation rates have
varied geographically throughout Earth's history (Peters 2005). At any given time in the past,
most of the habitable environments that existed are unpreserved. This facet of the rock record
becomes particularly problematic when a clade endemic to a region either appears or disappears
due to a shift in preservation potential. If ψ does not vary to accommodate the change in local
sampling rates, then the sudden change in diversity may be incorrectly inferred to reflect a
change in the diversification parameters.

Within-lineage preservation heterogeneity can also impact macroevolutionary inference,
and has been demonstrated to influence estimates of diversification (Liow et al. 2010b; Silvestro
et al. 2014). Factors such as geographic occupancy (Foote et al. 2007, 2016) and expected
abundance (Stanley 1979) are thought to vary with the age of a lineage, and both can influence

698 preservation potential. The probability of preservation may also vary among lineages within a
phylogeny. Many factors, from body size and shell composition to habitat preferences and
700 foraging strategies, influence the preservation potential of a taxon (Valentine 1989;
Behrensmeyer et al. 1992, 2000; Valentine et al. 2006). Furthermore, the magnitude and
702 directionality of these biases can vary with time and underlying sedimentology (Kidwell 2005;
Foote et al. 2015). The BAMM extension described in this article incorporates a very simple
704 model of preservation, such that users merely need to specify the total number of fossil
occurrences associated with the lineages that are included in the phylogeny. Accommodating
706 changes in preservation potential through time, among lineages, and across geographic regions is
an obvious future extension for models like BAMM that integrate extant and fossil data.

Comparability of Neontological and Paleontological Data

710 Many challenges remain in linking paleontological and neontological data, as the very
concept of a lineage differs between the two datatypes. Fossil taxa are identified based on
712 preserved morphological traits, such that that a single fossil morphospecies may be equivalent to
multiple biological species that are distinguishable by multiple traits that are unlikely to fossilize
714 (e.g., behavior, coloration, soft-tissue anatomy, vocalizations). Further, it is difficult in the fossil
record to discriminate between a single lineage sampled repeatedly through time (a series of
716 anagenetic morphospecies) and multiple, closely-related but separate lineages present in different
stratigraphic horizons. Hence, inferences that use method described in this article are subject to
718 an additional source of uncertainty that reflects different species concepts (either in theory or in
practice) between neontological and paleontological data.

How Many Fossil Occurrences?

The model we have implemented assumes that fossil occurrences are generated under a simple birth-death-fossilization process, which implies that occurrences represent the sampling of a specific lineage at a specific point in time. We suggest that the number of occurrences should reflect the number of temporally-unique occurrences of identifiable lineages, or the number of stratigraphically unique occurrences of both extinct and extant species-level taxa that are included in the phylogeny. The number of temporally-unique identifiable occurrences differs from raw counts of the total number of fossil occurrences of the clade in several important ways. In BAMM, taxonomically indeterminate records belonging to a clade (e.g., “Aves indet.”) should not be used, as there is no way to know whether the fossil corresponds to an observed lineage within the focal clade (either an extinct or extant form represented by a lineage on the phylogeny), or if it belongs to an as-yet unrecognized branch within the clade (e.g., a new species not yet included in the phylogeny). Assuming we are generally working with species-level phylogenies, one should also avoid using occurrences that have not been resolved below the genus level in BAMM, as they might represent lineages that are not present in the tree. Furthermore, multiple individuals of the same species collected from a single bone or shell bed should be treated as a single occurrence.

As an example of how researchers might specify an appropriate number of fossil occurrences for a BAMM analysis, we will consider how occurrences might be tabulated for a recent phylogenetic study of extant and extinct canid mammals (Matzke and Wright 2016). A simple search (February 17, 2017) on “Canidae” from the Paleobiology Database (Alroy et al. 2001; Varela et al. 2015) gave at least 10,660 fossil observations (unique records). Of this set, 2,946 represented stratigraphically unique occurrences, and 2,107 of those observations were

resolved to the species level. The dataset analysed by Matzke and Wright (2016) contains 142 species-level taxa, and we identified 1,837 stratigraphically-unique occurrences of those taxa in the database. For the simple (time-homogeneous) preservation model described in this article, we would specify 1,837 occurrences for a BAMM analysis of the canid phylogeny. The Paleobiology Database allows rapid searching for occurrences that match only the set of taxa present in a given phylogeny, and querying the output for duplicate localities can be done quickly (see Dryad scripts for our horse example).

At present, the model assumes that fossil occurrences are comparable through time and across clades. This assumption is clearly violated in practice, particularly when we consider large clades whose constituent species differ in key biological traits that influence their probability of preservation. To illustrate this point, consider isolated vertebrate teeth, a common type of fossil. Crocodylians shed their teeth continuously through life, although their teeth are rarely identifiable to the species level. On the other hand, mammals only have two sets of teeth in their life, so they produce fewer fossils, but each tooth is far more likely to be mapped to an observed lineage on a phylogeny. These differences make it difficult to equate the occurrence of a single crocodylian tooth to the occurrence of a mammalian molar. An obvious extension to the modeling framework presented here will be to incorporate lineage-specific variation in preservation rates, which will help account for such clade-specific differences in preservation potential.

CONCLUSIONS

The fossilized birth-death process has already shown great promise for inferring and time-scaling phylogenies using all available evidence. Recent advances in extending the model

to incorporate piece-wise time variation (Gavryushkina et al. 2014, 2017) and non-uniform
incomplete sampling (Zhang et al. 2016) have greatly improved the ability to estimate
phylogenies that include both extinct and extant lineages. Here, we have shown that BAMM
performs well on dated phylogenies that include a mix of extant and extinct tips (or extinct-only
tips), providing a useful tool for characterizing diversification rate variation across the expanding
pool of time-calibrated phylogenies that include extinct lineages. This approach will help
researchers test hypotheses about macroevolutionary rate control in clades that may lack
sufficient fossil data for non-phylogenetic (occurrence-only) paleobiological inferences.

The results of this study and others (e.g., Didier et al. 2012) demonstrate that the
inclusion of fossil data improves the estimation of both speciation and extinction rates from
phylogenies. The assumptions of the model we describe are simplistic, and numerous challenges
remain to be addressed. However, we have found that results are reasonably robust to several key
violations of model assumptions. The method described here is a first step towards integrating
information from paleontological and neontological datasets to better understand the extent to
which the dynamics of species diversification have varied through time and among lineages.

SUPPLEMENTARY MATERIAL

All simulation results and analysis code are available on Dryad, doi:10.5061/dryad.p009h.

BAMM v2.6 is also available on GitHub, at <https://github.com/macroevolution/bamm>.

FUNDING

This work was supported by the U. S. National Science Foundation (NSF-DEB-1256330 to
DLR), by a fellowship from the David and Lucile Packard Foundation (DLR), and by a VICI
grant (RSE) from the Netherlands Organization for Scientific Research (NWO).

ACKNOWLEDGMENTS

We thank R. Ree and three anonymous reviewers for comments on the manuscript.

REFERENCES

- Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G.,
Harmon L.J. 2009. Nine exceptional radiations plus high turnover explain species
diversity in jawed vertebrates. *Proc. Natl. Acad. Sci.* 106:13410–13414.
- Alroy J. 1999. The Fossil Record of North American Mammals: Evidence for a Paleocene
Evolutionary Radiation. *Syst. Biol.* 48:107–118.
- Alroy J. 2008. Dynamics of origination and extinction in the marine fossil record. *Proc. Natl.
Acad. Sci.* 105:11536–11542.
- Alroy, J. 2010. The shifting balance of diversity among major marine animal groups. *Science*
329: 1191 - 1194.
- Alroy J., Marshall C.R., Bambach R.K., Bezusko K., Foote M., Fürsich F.T., Hansen T.A.,
Holland S.M., Ivany L.C., Jablonski D., Jacobs D.K., Jones D.C., Kosnik M.A., Lidgard
S., Low S., Miller A.I., Novack-Gottshall P.M., Olszewski T.D., Patzkowsky M.E., Raup
D.M., Roy K., Sepkoski J.J., Sommers M.G., Wagner P.J., Webber A. 2001. Effects of
sampling standardization on estimates of Phanerozoic marine diversification. *Proc. Natl.
Acad. Sci.* 98:6261–6266.

- 812 Beaulieu J.M., O'Meara B.C. 2015. Extinction can be estimated from moderately sized
molecular phylogenies. *Evolution*. 69:1036–1043.
- 814 Behrensmeyer A.K., Hook R.W., Badgley C., Boy J., Chapman R., Dodson P., Gastaldo R.,
Graham R., Martin L., Olsen P. 1992. Paleoenvironmental contexts and taphonomic
816 modes. *Terr. Ecosyst. Time Evol. Paleoecol. Terr. Plants Anim. Univ. Chic. Press*
Chic.:15–136.
- 818 Behrensmeyer A.K., Kidwell S.M., Gastaldo R.A. 2000. Taphonomy and paleobiology.
Paleobiology. 26:103–147.
- 820 Benton M.J., Pearson P.N. 2001. Speciation in the fossil record. *Trends Ecol. Evol.* 16:405–411.
Cantalapiedra J.L., Prado J.L., Fernández M.H., Alberdi M.T. 2017. Decoupled
822 ecomorphological evolution and diversification in Neogene-Quaternary horses. *Science*.
355:627–630.
- 824 Carrano M.T., Oreska M.P.J., Lockwood R. 2016. Vertebrate paleontology of the Cloverly
Formation (Lower Cretaceous), II: Paleoecology. *J. Vertebr. Paleontol.* 36:e1071265.
- 826 Didier G., Fau M., Laurin M. 2015. Likelihood of tree topologies with fossils and diversification
rate estimation. *Systematic Biology* 6 : 964-987.
- 828 Didier G., Royer-Carenzi M., Laurin M. 2012. The reconstructed evolutionary process with the
fossil record. *J. Theor. Biol.* 315:26–37.
- 830 Etienne R.S., Haegeman B. 2012. A Conceptual and Statistical Framework for Adaptive
Radiations with a Key Role for Diversity Dependence. *Am. Nat.* 180:E75–E89.
- 832 Etienne R.S., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B. 2012.
Diversity-dependence brings molecular phylogenies closer to agreement with the fossil
834 record. *Proc. R. Soc. Lond. B Biol. Sci.* 279:1300–1309.

- Ezard T.H.G., Purvis A. 2016. Environmental changes define ecological limits to species richness and reveal the mode of macroevolutionary competition. *Ecol. Letters*. 19:899–906. doi:10.1111/ele.12626.
- FitzJohn R.G., Maddison W.P., Otto S.P. 2009. Estimating Trait-Dependent Speciation and Extinction Rates from Incompletely Resolved Phylogenies. *Syst. Biol.* 58:595–611.
- Foote M., Crampton J.S., Beu A.G., Marshall B.A., Cooper R.A., Maxwell P.A., Matcham I. 2007. Rise and fall of species occupancy in Cenozoic fossil mollusks. *Science*. 318:1131–1134.
- Foote M., Crampton J.S., Beu A.G., Nelson C.S. 2015. Aragonite bias, and lack of bias, in the fossil record: lithological, environmental, and ecological controls. *Paleobiology*. 41:245–265.
- Foote M., Raup D.M. 1996. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology*. 22:121–140.
- Foote M., Ritterbush K.A., Miller A.I. 2016. Geographic ranges of genera and their constituent species: structure, evolutionary dynamics, and extinction resistance. *Paleobiology*. 42:269–288.
- Gavryushkina A., Heath T.A., Ksepka D.T., Stadler T., Welch D., Drummond A.J. 2017. Bayesian Total-Evidence Dating Reveals the Recent Crown Radiation of Penguins. *Syst. Biol.* 66:57–73.
- Gavryushkina A., Welch D., Stadler T., Drummond A.J. 2014. Bayesian Inference of Sampled Ancestor Trees for Epidemiology and Fossil Calibration. *PLOS Comput. Biol.* 10:e1003919.

- Heath T.A., Huelsenbeck J.P., Stadler T. 2014. The fossilized birth–death process for coherent
858 calibration of divergence-time estimates. *Proc. Natl. Acad. Sci.* 111:E2957–E2966.
- Heath T.A., Huelsenbeck J.P., Stadler T. 2014. The fossilized birth–death process for coherent
860 calibration of divergence-time estimates. *Proc. Natl. Acad. Sci.* 111:E2957–E2966.
- Holland S.M. 2016. The non-uniformity of fossil preservation. *Phil Trans R Soc B.*
862 371:20150130.
- Hunt G., Slater G. 2016. Integrating Paleontological and Phylogenetic Approaches to
864 Macroevolution. *Annu. Rev. Ecol. Evol. Syst.* 47:null.
- Jablonski D., Shubin N.H. 2015. The future of the fossil record: Paleontology in the 21st century.
866 *Proc. Natl. Acad. Sci.* 112:4852–4858.
- Kidwell S.M. 2005. Shell Composition Has No Net Impact on Large-Scale Evolutionary Patterns
868 in Mollusks. *Science.* 307:914–917.
- Ksepka D.T., Parham J.F., Allman J.F., Benton M.J., Carrano M.T., Cranston K.A., Donoghue
870 P.C.J., Head J.J., Hermesen E.J., Irmis R.B., Joyce W.G., Kohli M., Lamm K.D., Leehr D.,
Patané J.L., Polly P.D., Phillips M.J., Smith N.A., Smith N.D., Tuinen M.V., Ware J.L.,
872 Warnock R.C.M. 2015. The Fossil Calibration Database—A New Resource for
Divergence Dating. *Syst. Biol.* 64:853–859.
- 874 Liow L.H., Quental T.B., Marshall C.R. 2010a. When Can Decreasing Diversification Rates Be
Detected with Molecular Phylogenies and the Fossil Record? *Syst. Biol.* 59:646–659.
- 876 Liow L.H., Skaug H.J., Ergon T., Schweder T. 2010b. Global occurrence trajectories of
microfossils: environmental volatility and the rise and fall of individual species.
878 *Paleobiology.* 36:224–252.

- Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a Binary Character's Effect on
880 Speciation and Extinction. *Syst. Biol.* 56:701–710.
- Matzke, N. J., Wright, A. 2016. Inferring node dates from tip dates in fossil Canidae: the
882 importance of tree priors. *Biol. Lett.* 12: 20160328
- Mitchell J.S., Makovicky P.J. 2014. Low ecological disparity in Early Cretaceous birds. *Proc. R.*
884 *Soc. Lond. B Biol. Sci.* 281:20140608.
- Mitchell J.S., Rabosky D.L. 2016. Bayesian model selection with Bayesian Analysis of
886 Macroevolutionary Mixtures: effects of the model prior on the inferred number of
diversification shifts. *Methods Ecol. Evol.* doi: 10.1111/2041-210X.12626.
- 888 Moore, B. R., Höhna, S., May, M. R., Rannalla, B., Huelsenbeck, J. P. 2016. Critically
evaluating the theory and performance of Bayesian analysis of macroevolutionary
890 mixtures. *Proc. Natl. Acad. Sci.* 113:9569-9574.
- Morlon, H. 2014. Phylogenetic approaches for studying diversification. *Ecology Letters* 17:508-
892 525.
- Morlon H., Parsons T.L., Plotkin J.B. 2011. Reconciling molecular phylogenies with the fossil
894 record. *Proc. Natl. Acad. Sci.* 108:16327–16332.
- Nee S. 2006. Birth-Death Models in Macroevolution. *Annu. Rev. Ecol. Evol. Syst.* 37:1–17.
- 896 Nee S., May R.M., Harvey P.H. 1994. The Reconstructed Evolutionary Process. *Philos. Trans. R.*
Soc. Lond. B Biol. Sci. 344:305–311.
- 898 Peters S.E. 2005. Geologic constraints on the macroevolutionary history of marine animals. *Proc.*
Natl. Acad. Sci. U. S. A. 102:12326–12331.
- 900 Peters S.E. 2006. Macrostratigraphy of North America. *J. Geol.* 114:391–412.

- Peters S.E., Foote M. 2001. Biodiversity in the Phanerozoic: a reinterpretation. *Paleobiology*.
902 27:583–601.
- Pujos F., Iuliis G.D., Cartelle C. 2016. A Paleogeographic Overview of Tropical Fossil Sloths:
904 Towards an Understanding of the Origin of Extant Suspensory Sloths? *J. Mamm. Evol.*:1–20.
- Purvis A. 2008. Phylogenetic Approaches to the Study of Extinction. *Annu. Rev. Ecol. Evol. Syst.* 39:301–319.
- 908 Quental T.B. Marshall C.R. 2010. Diversity dynamics: molecular phylogenies need the fossil record. *Trends Ecol. Evol.* 25:434–441.
- 910 Quental, T.B., Marshall C. R. 2013. How the Red Queen drives terrestrial mammals to extinction. *Science* 341:290 - 292.
- 912 Rabosky D.L. 2009. Heritability of Extinction Rates Links Diversification Patterns in Molecular Phylogenies and Fossils. *Syst. Biol.* 58:629–640.
- 914 Rabosky D.L. 2010. Extinction Rates Should Not Be Estimated from Molecular Phylogenies. *Evolution*. 64:1816–1824.
- 916 Rabosky D.L. 2014. Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. *PLOS ONE*. 9:e89543.
- 918 Rabosky D.L. 2016. Challenges in the estimation of extinction from molecular phylogenies: A response to Beaulieu and O’Meara. *Evolution*. 70:218–228.
- 920 Rabosky D.L., Grundler M., Anderson C., Title P., Shi J.J., Brown J.W., Huang H., Larson J.G. 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on
922 phylogenetic trees. *Methods Ecol. Evol.* 5:701–707.

- Rabosky, D. L., Mitchell, J. S., Chang J. 2017. Is BAMM flawed? Theoretical and practical
924 concerns in the analysis of multi-rate diversification models. *Syst. Biol.*
10.1093/sysbio/syx037.
- 926 Rabosky D.L., Santini F., Eastman J., Smith S.A., Sidlauskas B., Chang J., Alfaro M.E. 2013.
Rates of speciation and morphological evolution are correlated across the largest
928 vertebrate radiation. *Nat. Commun.* 4:1958.
- Raup D.M. 1972. Taxonomic Diversity during the Phanerozoic. *Science*. 177:1065–1071.
- 930 Raup D.M. 1976. Species diversity in the Phanerozoic: an interpretation. *Paleobiology*. 2:289–
297.
- 932 Raup D.M. 1985. Mathematical models of cladogenesis. *Paleobiology*. 11:42–52.
- Raup D.M. 1994. The role of extinction in evolution. *Proc. Natl. Acad. Sci.* 91:6758–6763.
- 934 Raup D.M., Gould S.J., Schopf T.J.M., Simberloff D.S. 1973. Stochastic Models of Phylogeny
and the Evolution of Diversity. *J. Geol.* 81:525–542.
- 936 Ronquist F., Klopstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn A.P. 2012. A
Total-Evidence Approach to Dating with Fossils, Applied to the Early Radiation of the
938 Hymenoptera. *Syst. Biol.* 61:973–999.
- Rosenblum E.B., Sarver B.A.J., Brown J.W., Roches S.D., Hardwick K.M., Hether T.D.,
940 Eastman J.M., Pennell M.W., Harmon L.J. 2012. Goldilocks Meets Santa Rosalia: An
Ephemeral Speciation Model Explains Patterns of Diversification Across Time Scales.
942 *Evol. Biol.* 39:255–261.
- Sakamoto, M., Benton, M. J., and C. Venditti. 2016. Dinosaurs in decline tens of millions of
944 years before their final extinction. *Proc. Natl. Acad. Sci. U.S.A.* 113:5036-5040.

- Sepkoski J.J. 1978. A Kinetic Model of Phanerozoic Taxonomic Diversity I. Analysis of Marine
946 Orders. *Paleobiology*. 4:223–251.
- Sepkoski J.J., Bambach R.K., Raup D.M., Valentine J.W. 1981. Phanerozoic marine diversity
948 and the fossil record. *Nature*. 293:435–437.
- Sheets H.D., Mitchell C.E., Melchin M.J., Loxton J., Štorch P., Carlucci K.L., Hawkins A.D.
950 2016. Graptolite community responses to global climate change and the Late Ordovician
mass extinction. *Proc. Natl. Acad. Sci.*:201602102.
- 952 Silvestro D., Schnitzler J., Liow L.H., Antonelli A., Salamin N. 2014. Bayesian Estimation of
Speciation and Extinction from Incomplete Fossil Occurrence Data. *Syst. Biol.* 63:349–
954 367.
- Silvestro D., Salamin N., Schnitzler J. 2014. PyRate: a new program to estimate speciation and
956 extinction rates from incomplete fossil data. *Methods Ecol. Evol.* 5:1126–1131.
- Stadler T. 2010. Sampling-through-time in birth–death trees. *J. Theor. Biol.* 267:396–404.
- 958 Stadler T. 2013. Recovering speciation and extinction dynamics based on phylogenies. *J. Evol.*
Biol. 26:1203–1219.
- 960 Stanley S.M. 1979. *Macroevolution, Pattern and Process*. Johns Hopkins University Press.
- Stock C. 1929. A Census of the Pleistocene Mammals of Rancho La Brea, Based on the
962 Collections of the Los Angeles Museum. *J. Mammal.* 10:281–289.
- Valentine J.W. 1989. How good was the fossil record? Clues from the Californian Pleistocene.
964 *Paleobiology*. 15:83–94.
- Valentine J.W., Jablonski D., Kidwell S., Roy K. 2006. Assessing the fidelity of the fossil record
966 by using marine bivalves. *Proc. Natl. Acad. Sci. U. S. A.* 103:6599–6604.

Varela S., González-Hernández J., Sgarbi L.F., Marshall C., Uhen M.D., Peters S., McClennen

968 M. 2015. paleobioDB: an R package for downloading, visualizing and processing data
from the Paleobiology Database. *Ecography*. 38:419–425.

970 Zanno L.E., Makovicky P.J. 2011. Herbivorous ecomorphology and specialization patterns in
theropod dinosaur evolution. *Proc. Natl. Acad. Sci.* 108:232–237.

972 Zhang C., Stadler T., Klopstein S., Heath T.A., Ronquist F. 2016. Total-Evidence Dating under
the Fossilized Birth–Death Process. *Syst. Biol.* 65:228–249.

974

976 Figure Legends

Figure 1. Diagram showing a complete evolutionary tree with unobserved (gray) and observed (black) evolutionary history. Fossils are indicated by open circles; labeled circles (a, b) denote rate shifts on individual branches. Tree contains three extant tips, three observed extinct tips, and eight unobserved extinct tips. Extinct taxa are necessarily placed at the time of their last fossil occurrence. Rate shift (a) is unobserved in phylogenies that include fossil data, as there are no extant descendants or fossil occurrences on the subtree descended from the shift. The simulation protocol used in this study involved generating phylogenies for the complete process, then simulating fossil occurrences on the tree and removing unobserved (gray) evolutionary history.

Figure 2. BAMM estimates of speciation (λ), extinction (μ) and relative extinction rate (μ / λ) estimates as a function of the true values for rate regimes with at least 5 tips from trees fossilized with different preservation rates. Each gray point is the estimated rate for a single rate regime. Black points represent mean estimated and inferred rates for all regimes falling within each 2% quantile of true rates. For example, the mean estimated rate of all individual regimes with a true rate between the lowest and the 2% quantile comprise the first black point. Rows give results for extant-only trees (first row), low ($\psi = 0.01$; second row), moderate ($\psi = 0.1$; third row) and high ($\psi = 1.0$; fourth row). Trees with higher preservation rates, and thus more fossils, have better rate estimates, with the extinction rate showing particular improvement.

Figure 3. BAMM estimates of speciation (λ), extinction (μ) and relative extinction (μ / λ) rates for rate regimes with at least 20 observed tips from trees fossilized with different preservation

998 rates (ψ). Each gray point represents the mean posterior estimate of rates for a regime from the
extant-only trees (first row), low ($\psi = 0.01$; second row), moderate ($\psi = 0.1$; third row) and high
1000 ($\psi = 1.0$; fourth row). Each black point represents the value of all regimes falling within each 2%
quantile (e.g., all regimes between the 0th and 2nd percent quantile for the first point). Rates in
1002 general are better estimated in the large regimes (as compared to Fig. 2). Increased preservation
rates lead to better macroevolutionary rate estimation overall, particularly for extinction.

1004
Figure 4. Speciation (λ) and extinction (μ) rate correlations across all regimes as a function of the
1006 minimum number of tips present in a particular subset of the data. Each point represents the
correlation for speciation (a) or extinction (b) for all rate regimes with the specified minimum
1008 number of tips. A minimum regime size (x-axis) value of zero indicates that the correlation is
computed across all rate regimes, regardless of the number of tips they contain; a minimum value
1010 of 10 would restrict this pool of regimes to only those with 10 or more tips. Minimum values of 5
and 20 correspond to results shown in figures 2 and 3, respectively. Larger regimes yield better
1012 rate estimates for both speciation and extinction. Speciation rates are reliably estimated even in
the absence of fossil data, but extinction rate accuracy is markedly improved by the inclusion of
1014 fossils.

1016 Figure 5. Marginal posterior distributions for the preservation rate (ψ) for trees simulated with
preservation rates of 0.01, 0.1 and 1.0. Black arrows denote the true value of ψ , while gray bars
1018 show the distribution of mean ψ across all 200 trees for each value.

1020 Figure 6. The true number of shifts in the degraded trees plotted against the mean number of
1022 inferred shifts (number of non-root rate regimes best supported by stepwise model selection
1024 using Bayes factors) for preservation rates (ψ) of 0.01, 0.1 and 1.0. Each point represents one of
the 200 trees. The identity line is shown for reference, and the estimates slopes are very low
(0.08, 0.085 and 0.18). The correlation between true and estimated numbers of shifts are 0.65,
0.73 and 0.8 for the preservations rates from lowest to highest. Plots in the top row show the
results on the identical axes, while the plots in the second row show the same data with the y-
axis rescaled to better show the correlation between the true and estimated number of shifts.
Most simulated shifts are small in terms of the number of lineages and they may also reflect a
relatively minor change in rates (e.g., a shift from speciation of 0.1 to 0.11). As such, a method
focused on summarizing the major regimes of the tree (i.e., those with strong support) is
expected to underestimate the number of shifts. This is especially true when a stringent Bayes
factor threshold is used (see SOM for the mean number of shifts from the posterior distribution).

1034 Figure 7. Topological accuracy of inferred shift locations. (a) Marginal posterior probability of
each rate shift as a function of the number of tips in the shift regime. A value of 0.50 would
imply that 50% of samples from the posterior recovered a shift in the same topological location
(e.g., same branch) as the true shift. In general, we expect the marginal shift probability to be
somewhat less than 1.0 even for well-supported shifts, because the evidence for a given shift can
be smeared across several adjacent branches (Rabosky et al. 2014). Red horizontal lines denote
median marginal probabilities of rate shifts within sliding windows of arbitrary size (see text for
details). (b) Marginal probability of each rate shift as a function of both the proportional change
in speciation rate ($\lambda_{\text{NEW}} / \lambda_{\text{OLD}}$) as well as the number of descendant tips in the regime. Darker

points represent higher marginal probabilities. Shift locations are inferred with the highest accuracy is highest when large rate increases occur and/or when shifts lead to large numbers of descendant tips (c) Accuracy with which BAMM infers cohort membership, computed as the mean probability of assigning a given taxon pair to the correct grouping relationship (e.g. "same shift regime", "different shift regime"). Red points represent the mean cohort accuracy of BAMM analyses across four preservation rates as a function of the number of shifts. Black squares represent the corresponding cohort assignment accuracy expected under random shift placement; see text for details.

Figure 8. Performance of BAMM when the assumption of time-homogeneous fossil preservation is violated. Speciation (λ ; a), relative extinction (μ / λ ; b), inferred number of shifts (c) and cohort accuracy (d) are shown for trees fossilized using a stage-specific simulation and analyzed using the constant-preservation rate assumption of BAMM; results illustrate regimes with 5 or more tips, for comparison to Figure 2.

Figure 9. Bias in the estimation of macroevolutionary rates when the number of fossil occurrences was misspecified. The minimum number of fossil occurrences (F_{MIN}) is equal to the number of observed extinct tips, and the gap between the minimum and true number of fossil occurrences (F_{GAP}) was used to analyzed different trees at comparable levels of misspecification. The bias in speciation (a) and extinction (b) show a similar pattern of inflated rate estimates when fewer than the true number of observed fossils are specified, and deflated estimates when the number of fossils were exaggerated. Bold points represent the median bias, and red lines indicated the true number of fossils (vertical) and no bias in rate estimation (horizontal). The

1066 degrees of fossil misspecification are: 1) F_{MIN} , 2) $F_{\text{MIN}} + 0.25 \times F_{\text{GAP}}$, 3) $F_{\text{MIN}} + 0.5 \times F_{\text{GAP}}$, 4)
1068 F_{TRUE} , 5) $F_{\text{TRUE}} + F_{\text{GAP}}$, and 6) $10 \times F_{\text{TRUE}}$. Trees were simulated using constant rates through
time, and the inferred values are for the entire tree.

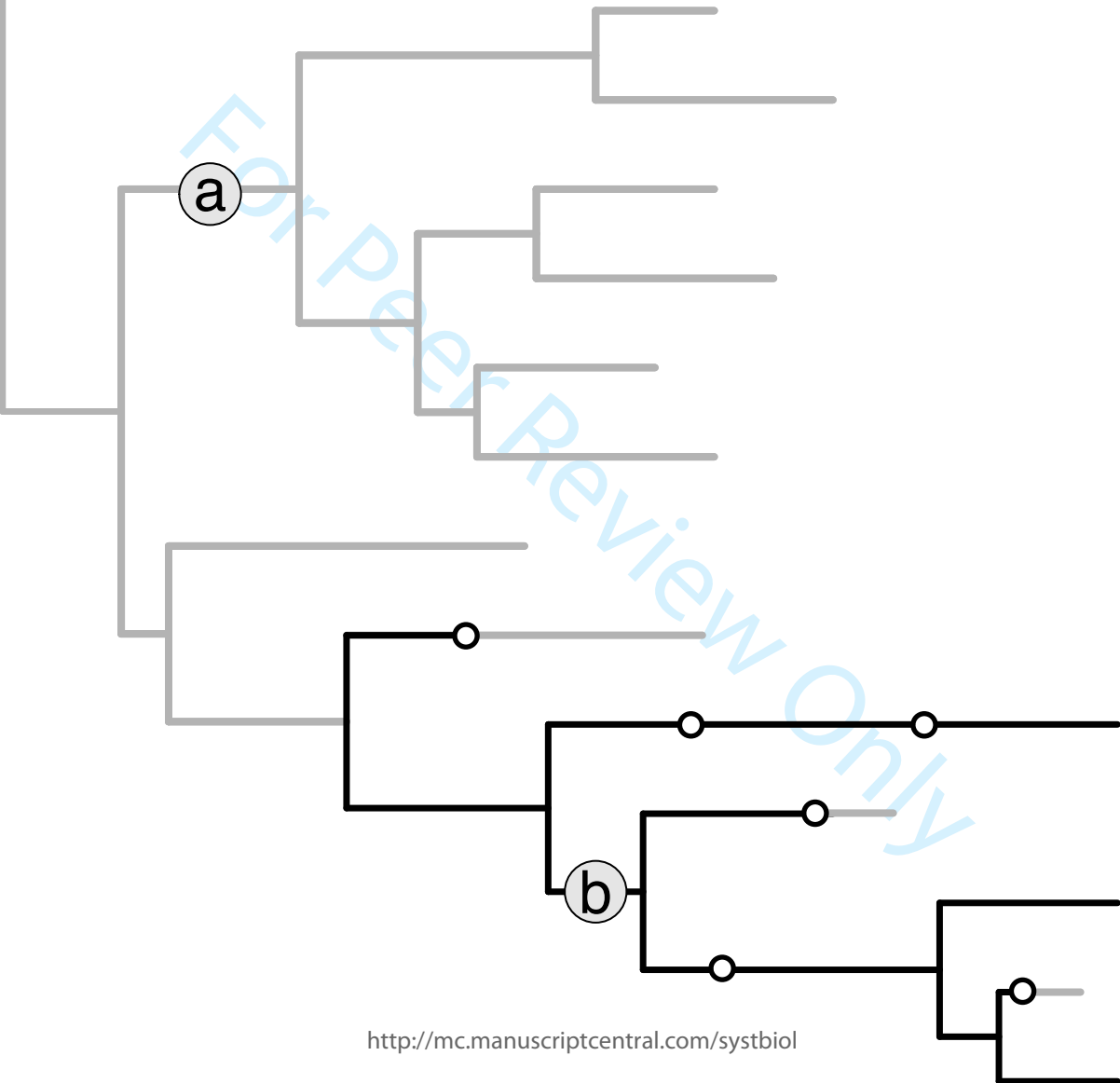
1070 Figure 10. Metrics of within-tree rate accuracy as a function of the proportion of shifts in the tree
that are unobserved, for trees with at least 50 observed tips. For the branches in each tree we
1072 evaluated the inferred and generating rate of speciation and extinction, calculated the slope (first
and third rows), and Pearson correlation (second and fourth rows) between them, and show them
1074 relative to the proportion of rate shifts in the tree that have no sampled taxa (see Fig. 1). The
relative density of values across all preservation levels is shown to the right. Both metrics for
1076 both rates have their highest density across all trees and rate regimes near the optimal value of 1
(mean slopes of 0.67 and 0.41 for speciation and extinction; mean correlations of 0.77 and 0.48
1078 for speciation and extinction). Neither rate shows a relationship between the within-tree metric
and the proportion of unobserved shifts suggesting that unobserved shifts do not bias rate
1080 estimates in these simulations. Filled circles denote trees where BAMM identified substantial
evidence for rate variation (Bayes factor evidence > 20 for one or more shifts); open circles are
1082 trees that lacked evidence for rate variation. Note that branch-specific slopes and correlations as
presented above are expected to equal zero if BAMM fails to detect diversification rate variation
1084 (see Rabosky et al. 2017, Fig. 10). For additional visualizations demonstrating the lack of a
relationship between performance and unobserved shifts, see the Supplementary Materials.

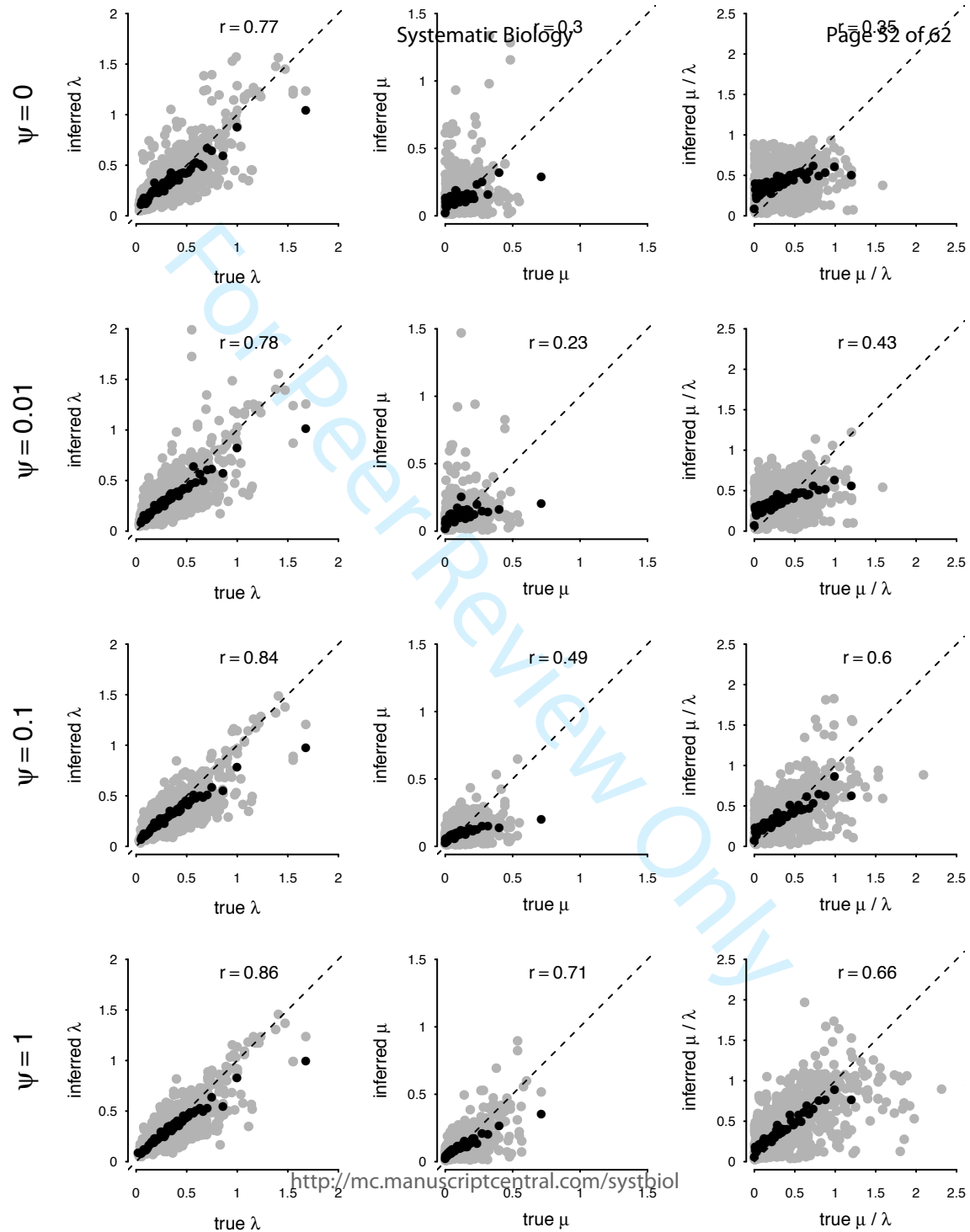
1086
Figure 11. Mean net diversification rates per branch (left), and mean speciation and extinction
1088 rates through time from BAMM and PyRate (right) for the horse phylogeny. PyRate and BAMM

estimates disagree with respect to the overall turnover and net diversification rates through time,
1090 although estimates of speciation are in accordance at the beginning of the horse phylogeny (~20-
15 Ma) and extinction estimates are comparable towards the present (~3-0 Ma).

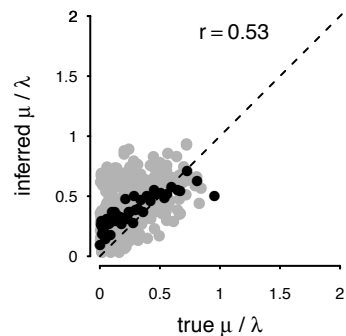
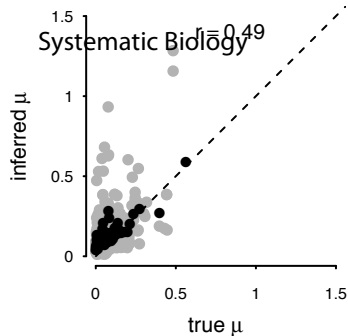
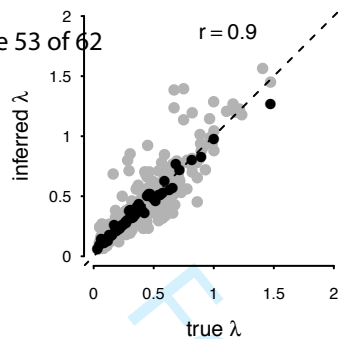
1092

For Peer Review Only

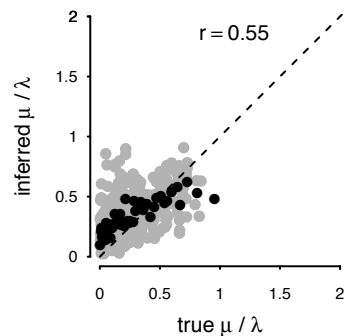
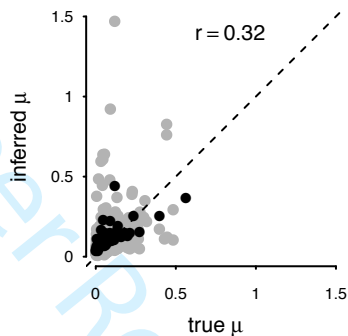
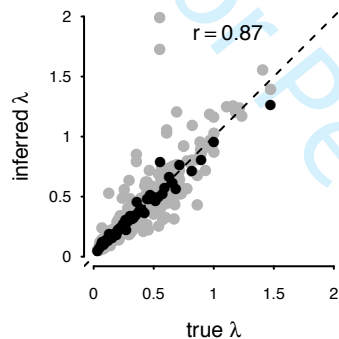




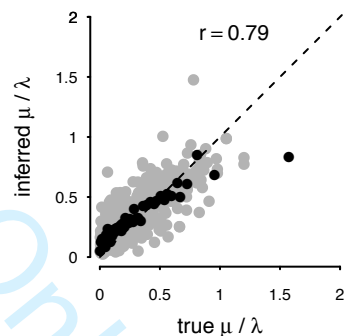
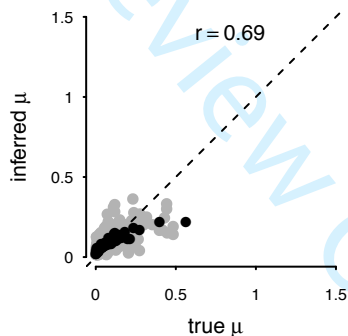
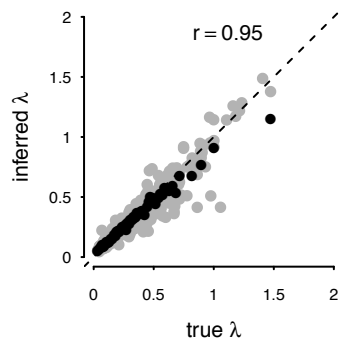
$\psi = 0$



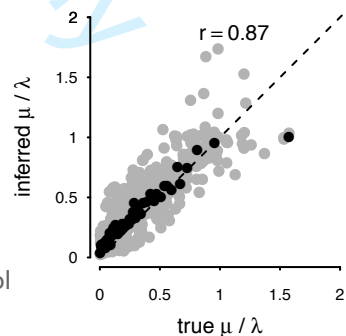
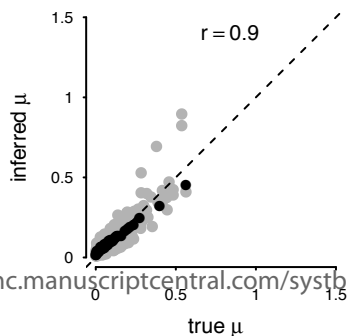
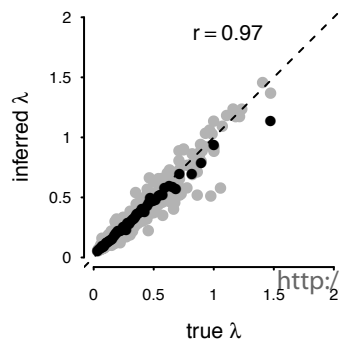
$\psi = 0.01$

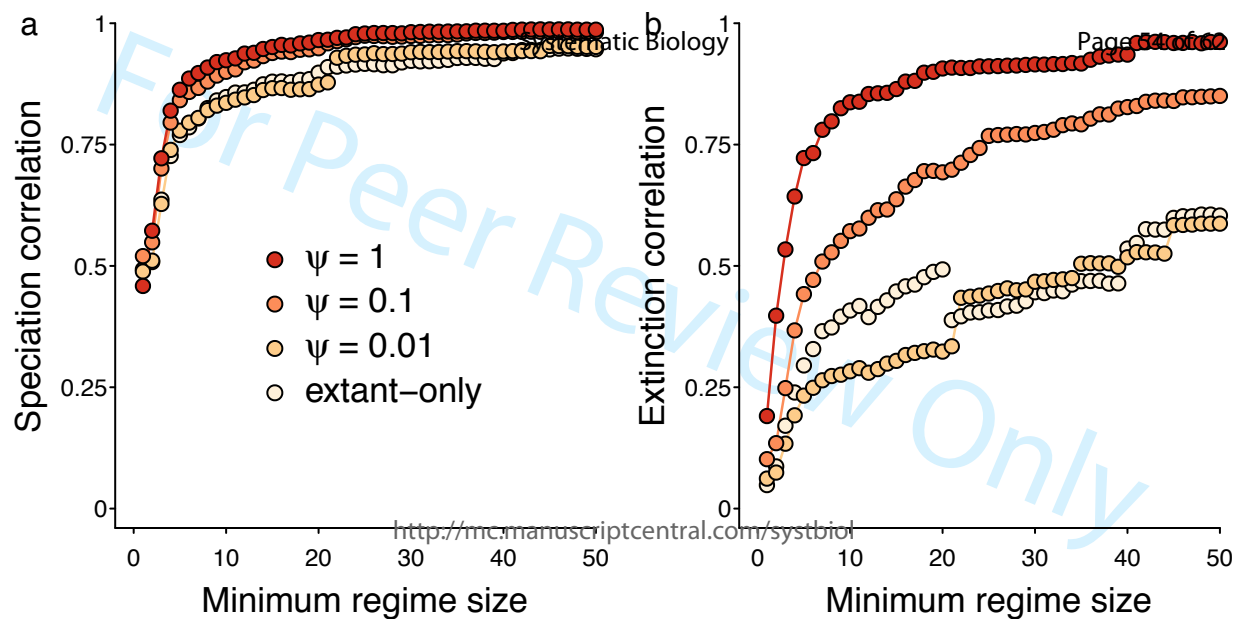


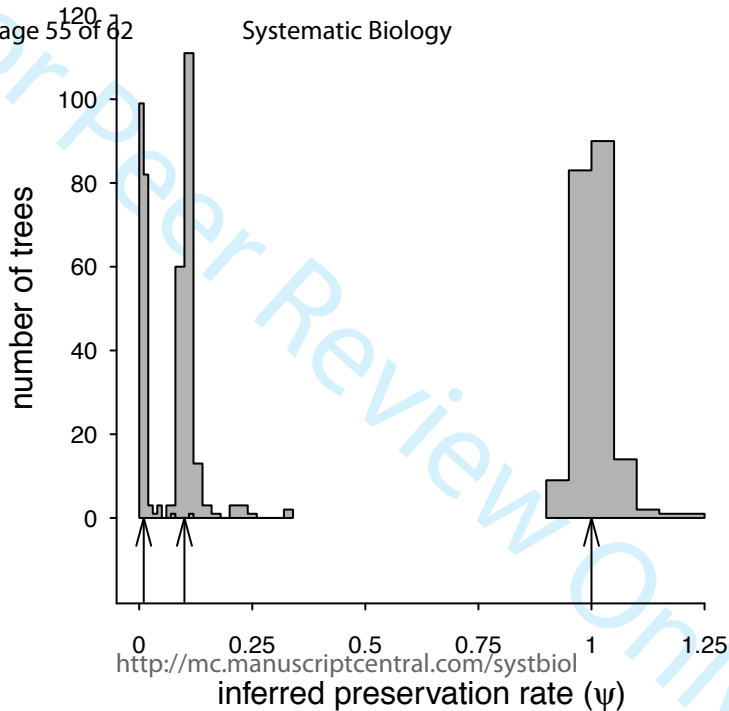
$\psi = 0.1$



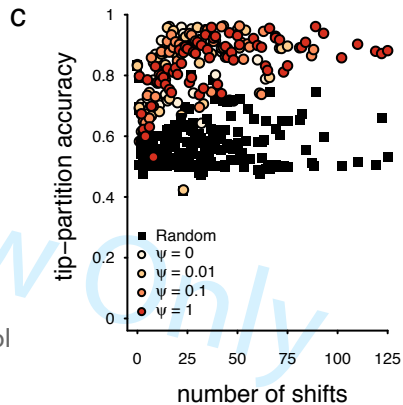
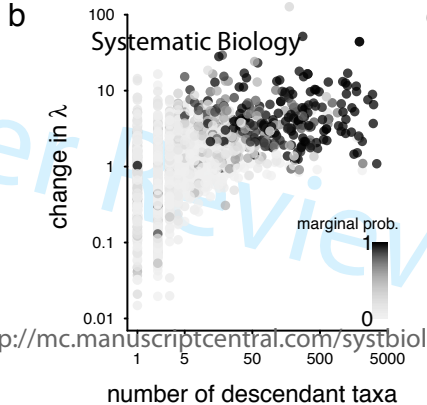
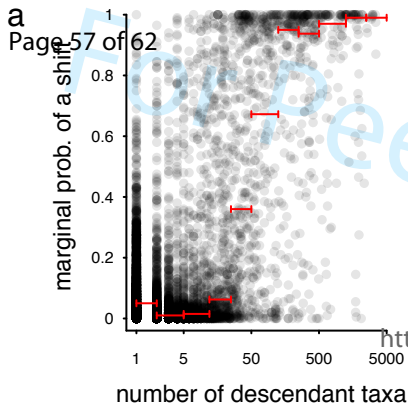
$\psi = 1$

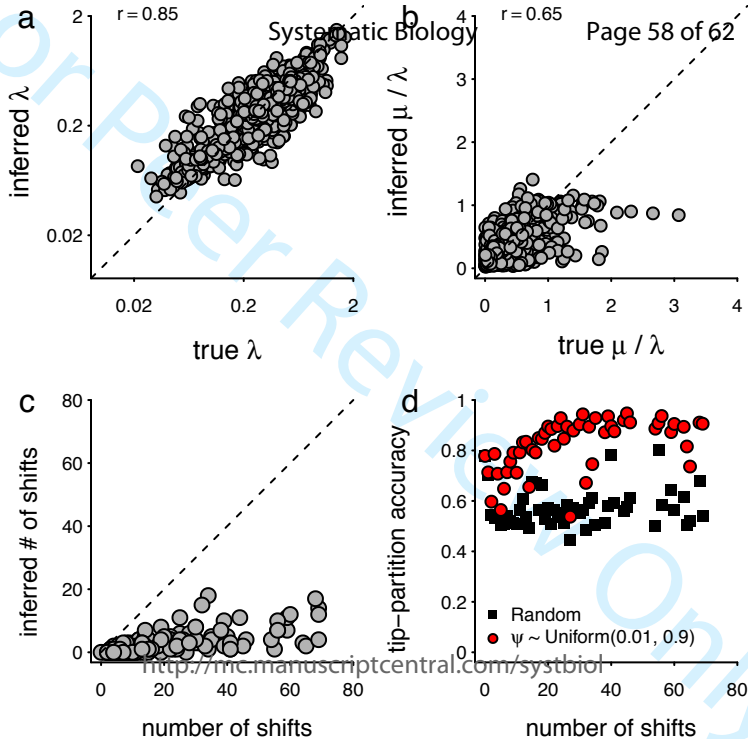




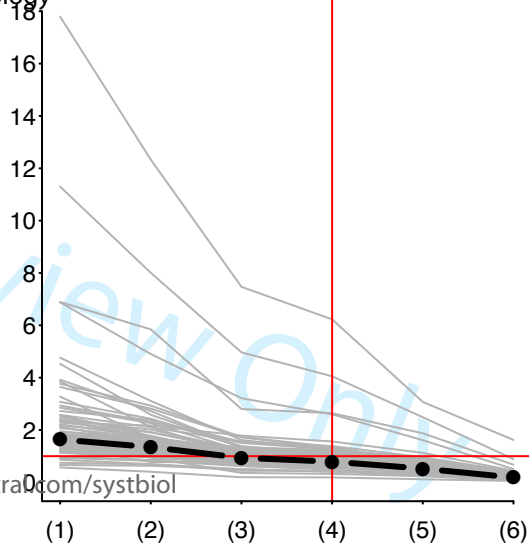
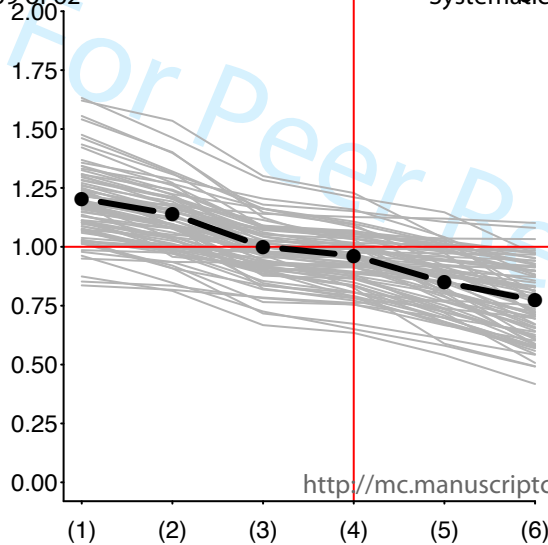


$\psi = 0.01$ $\psi = 0.1$ 





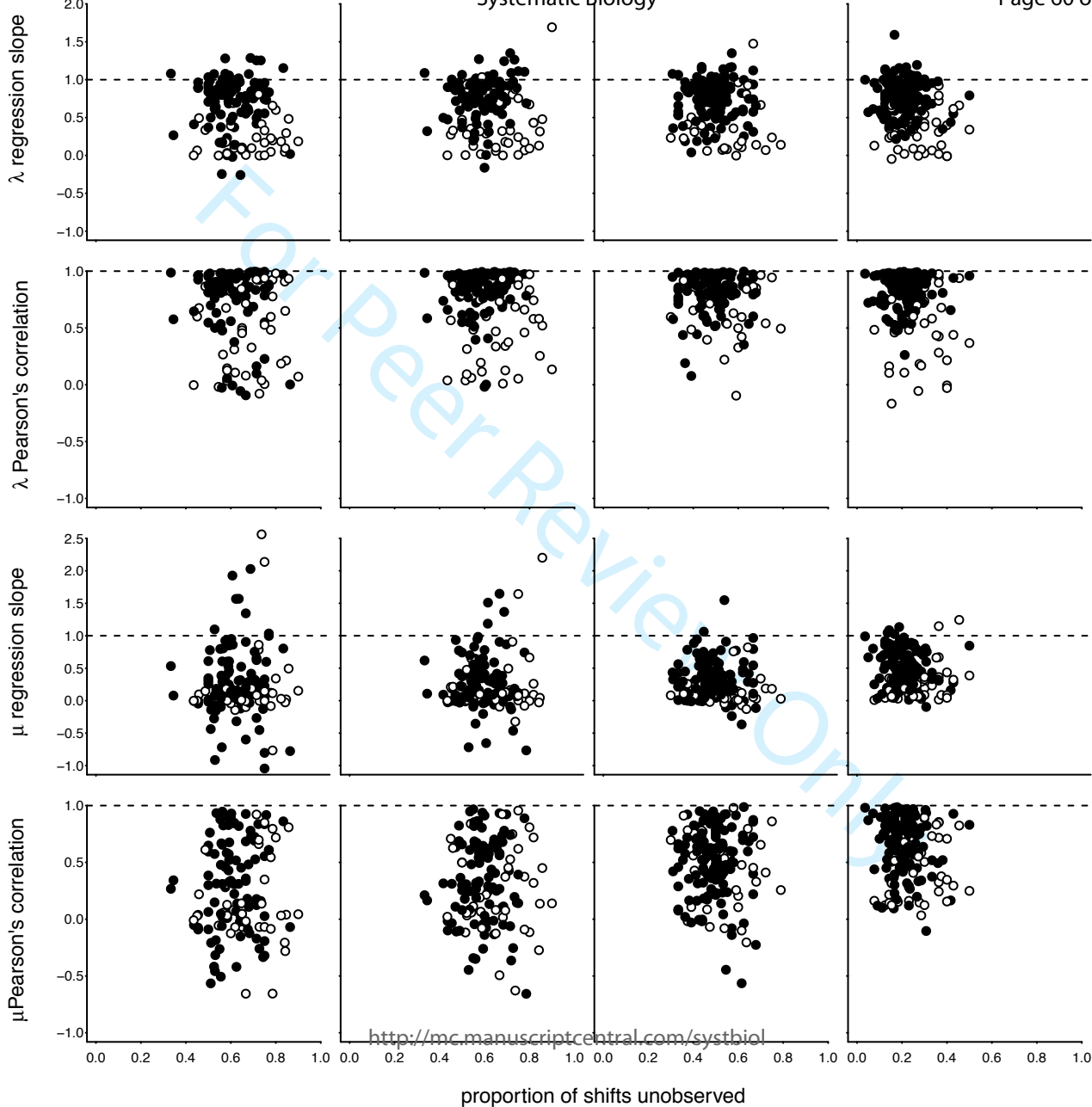
inferred rate / true rate

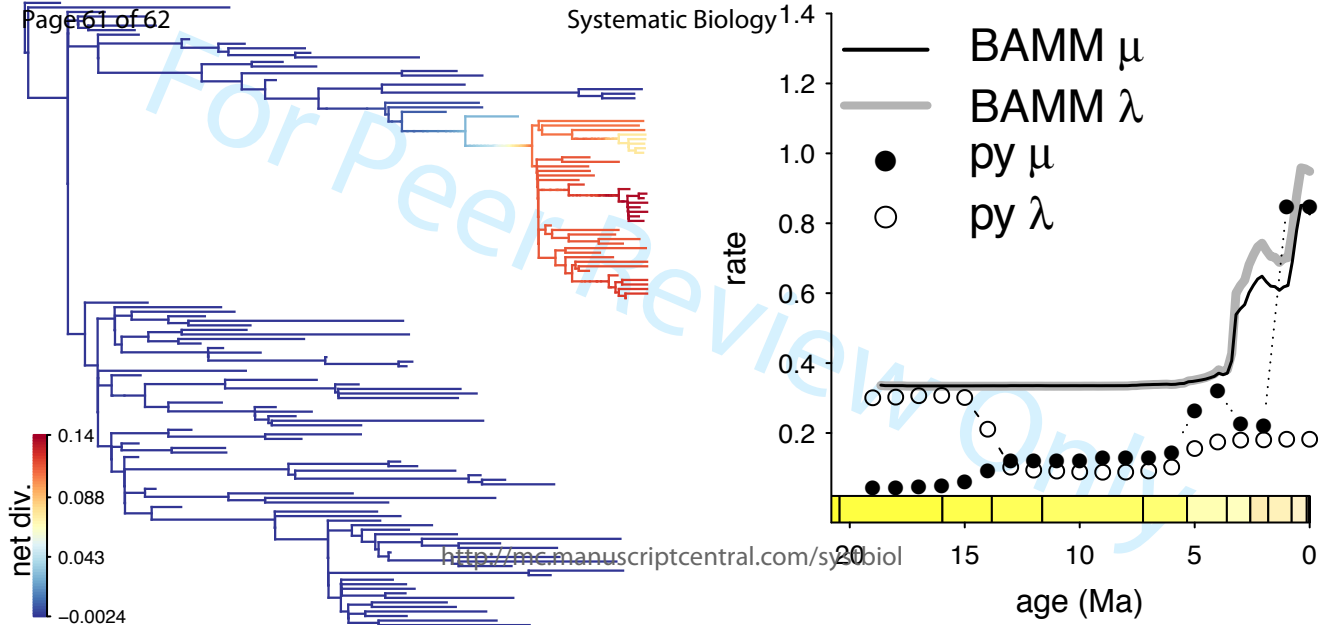


$\psi = 0$ $\psi = 0.01$ $\psi = 0.1$ $\psi = 1$

Systematic Biology

Page 60 of 62





Data file(s):
Supplementary Files and Data

Thank you for submitting your data package to Dryad for journal review. Please read the following information carefully and save it for future reference.

(1) Your data package is temporarily available to journal editors and reviewers via a private review link (see below).

(2) It has been assigned a Digital Object Identifier (DOI), but this DOI will not work until/unless your manuscript is accepted for publication and your data package is approved by our curators.

(3) If/when your manuscript is accepted, please make sure that the DOI, and NOT the review link, is included in the published article. More information about citing your data is below.

(4) If you need to modify your submission, please see the instructions below. Do not create multiple Dryad submissions associated with the same manuscript.

(5) If you were asked to enter a credit card number, your card will not be charged unless/until your manuscript is accepted and your submission is approved. You will receive a confirmation when your card is charged.

YOUR TEMPORARY REVIEW LINK:

<http://datadryad.org/review?doi=doi:10.5061/dryad.p009h>

(for journal access prior to manuscript acceptance only, DO NOT link in your manuscript)

YOUR DRYAD DOI:

doi:10.5061/dryad.p009h

(for inclusion in the published article)